Appendix B
Wenatchee Subbasin Plan

MONITORING STRATEGY FOR THE WENATCHEE RIVER BASIN

Draft Report

June 20, 2003

Prepared by:
Tracy W. Hillman
BioAnalysts, Inc.
Eagle, Idaho

Prepared for:
Upper Columbia Regional Technical Team
Upper Columbia Salmon Recovery Board
Wenatchee, Washington

TABLE OF CONTENTS

SECTION 1: INTRODUCTION

Managers often implement actions within tributary streams to improve the status of fish populations and their habitats. Until recently, there was little incentive to monitor such actions to see if they met their desired effects. Now, however, many programs require that funded actions include monitoring efforts. Within the Wenatchee Basin, Washington, several different organizations, including federal, state, tribal, local, and private entities currently implement tributary actions and conduct monitoring studies. Because of different goals and objectives, different entities are using different monitoring approaches and protocols. In some cases, however, different entities are measuring the same (or similar) things in the same streams with little coordination or awareness of each others efforts. The Upper Columbia Regional Technical Team (RTT) is aware of this problem and desires a monitoring strategy or plan that reduces redundancy, increases efficiency, and meets the goals and objectives of the various entities.

At least three different groups within the region have drafted integrated monitoring strategies that address many of the concerns of the RTT. For example, the Independent Scientific Advisory Board (ISAB) of the Northwest Power Planning Council outlined a monitoring and evaluation plan for assessing recovery of tributary habitat (ISAB 2003). They describe a three-tiered monitoring program that includes trend or routine monitoring (Tier 1), statistical (status) monitoring (Tier 2), and experimental research (effectiveness) monitoring (Tier 3). Trend monitoring obtains repeated measurements, usually representing a single spatial unit over a period of time, with a view to quantifying changes over time. Changes must be distinguished from background noise. This type of monitoring does not establish cause-and-effect relationships and does not provide inductive inferences to larger areas or time periods. Statistical monitoring, on the other hand, provides statistical inferences that extend to larger areas and longer time periods than the sample. This type of monitoring requires probabilistic selection of study sites and repeated visits over time. Experimental research monitoring is often required to establish cause-and-effect relationships between management actions and population/habitat response. This requires the use of experimental designs incorporating "treatments" and "controls" randomly assigned to study sites.

According to the ISAB (2003), the value of monitoring is greatly enhanced if the different types of monitoring are integrated. For example, trend and statistical monitoring will help define the issues that should be addressed with more intensive, experimental research monitoring. The latter will identify which habitat attributes are most informative and will provide conclusive information about the efficacy of various restoration approaches. Implementing experimental research in the absence of trend and statistical monitoring would increase uncertainty about the generalization of results beyond the sampling locations. The ISAB (2003) identified the following essential elements of a valid monitoring program.

- Develop a trend monitoring program based on remotely-sensed data obtained from sources such as aerial photography or satellite imagery or both.

- Develop and implement a long-term statistical monitoring program to evaluate the status of fish populations and habitat. This requires probabilistic (statistical) site selection procedures and establishment of common (standard) protocols and data collection methods.

- Implement experimental research monitoring at selected locations to establish the underlying causes for the changes in habitat and population indicators.

Another strategy developed by the Bonneville Power Administration, the U.S. Army Corps of Engineers, the Bureau of Reclamation (collectively referred to as the Action Agencies), and NOAA Fisheries responds to the Federal Columbia River Power System (FCRPS) Biological Opinion issued by the U.S. Fish and Wildlife Service and the National Marine Fisheries Service. Although the Action Agencies/NOAA Fisheries Draft Research, Monitoring, and Evaluation (RME) Program was developed before the release of the ISAB (2003) report, it is in many respects consistent with ISAB recommendations. For example, the draft RME Program calls for the classification of all watersheds that have listed fish populations and receive restoration actions. Classification is hierarchical and captures physical/environmental differences spanning from the largest scale (regional setting) down to the channel segment. This component of the draft RME Program comports with Tier 1 Trend Monitoring. Status Monitoring (similar to Tier 2 Statistical Monitoring) and Action Effectiveness Research (similar to Tier 3 Experimental Research) are also included in the RME Program. The ISAB is currently reviewing the RME Program.

About the time the Action Agencies/NOAA Fisheries released their draft program, the Washington Salmon Recovery Funding Board (WSRFB) released a draft monitoring and evaluation strategy for habitat restoration and acquisition projects. The document identified implementation, effectiveness, and validation monitoring as key components of their program. The monitoring program is scaled to capture factors operating at different hierarchical levels. At the lowest level (Level 0), the program determines if the action was implemented (implementation monitoring). Level 1 monitoring determines if projects meet the specified engineering and design criteria. Level 2 and 3 monitoring assess the effectiveness of projects on habitat and fish abundance, respectively. Levels 1-3 constitute effectiveness monitoring. Finally, level 4 (validation) monitoring addresses how management and habitat restoration actions, and their cumulative effects, affect fish production within a watershed. This type of monitoring is the most complex and technically rigorous.

Although the three programs (ISAB, Action Agencies/NOAA Fisheries, and WSRFB) describe monitoring in slightly different terms, they all address the same goal. That is, all three intend to assess the effectiveness of restoration projects and management actions on tributary habitat and fish populations. Consequently, the overall approaches among the three programs are similar, with the Action Agencies/NOAA Fisheries RME Program being the most intensive and extensive, in part because of the requirements of the FCRPS Biological Opinion. Indeed, the Action Agencies/NOAA Fisheries Program calls for monitoring all tributary actions with intensive, standardized protocols and data collection methods. For each tributary action, a list of specific indicators, ranging from water quality to watershed condition, are to be measured.

As noted earlier, various entities, including the Washington Salmon Recovery Fund Board, will be funding and implementing various restoration projects and actions within the Wenatchee Basin. These projects will be monitored to assess their effectiveness. Other groups, such as the U.S. Forest Service, U.S. Fish and Wildlife Service, Washington Department of Ecology, Washington Department of Fish and Wildlife, Chelan County, and Chelan County Public Utility District, will continue their ongoing monitoring of fish and habitat in the basin. In addition, NOAA Fisheries, with funding from the Bonneville Power Administration, will implement the status/trend monitoring component of the Action Agencies/NOAA Fisheries RME Plan in the basin. Because of all the activities occurring within the basin, it is important that the monitoring plan capture the needs of all entities, avoids duplication of sampling efforts, increases monitoring efficiency, and reduces overall monitoring costs.

The monitoring plan described in this document is not another regional monitoring strategy. Rather, this plan draws from the existing strategies (ISAB, Action Agencies/NOAA Fisheries, and WSRFB) and outlines an approach specific to the Wenatchee Basin. The plan described here addresses the following basic questions:

1. What are the current habitat conditions and abundance, distribution, life-stage survival, and age-composition of ESA-listed fish in the Wenatchee Basin (status monitoring)?
2. How do these factors change over time (trend monitoring)?
3. What effects do tributary habitat actions have on fish populations and habitat conditions (effectiveness monitoring)?

The plan is designed to address these questions and at the same time eliminate duplication of work, reduce costs, and increase monitoring efficiency. The implementation of valid statistical designs, probabilistic sampling designs, standardized data collection protocols, consistent data reporting methods, and selection of sensitive indicators will increase monitoring efficiency.[1] For this plan to be successful, all organizations involved must be willing to cooperate and freely share information. Cooperation includes sharing monitoring responsibilities, adjusting or changing sampling methods to comport with standardized protocols, and adhering to statistical design criteria. In those cases where the standardized method for measuring an indicator is different from what was used in the past, it may be necessary to measure the indicator with both methods for a few years so that a relationship can be developed between the two methods. Scores generated with a former method could then be adjusted to correct for any bias.

For convenience, I divided this report into eight major parts. The first part (Section 2) identifies valid statistical designs for status/trend and effectiveness monitoring. Section 3 discusses issues associated with sampling design, emphasizing how one selects a sample and how to minimize measurement error. Section 4 examines how sampling should occur at different spatial scales. Section 5 describes the importance of classification and identifies a suite of classification variables. Section 6 identifies and describes biological and physical/environmental indicators, while Section 7 identifies methods for measuring each indicator variable. These six sections provide the foundation for implementing an efficient monitoring plan in the Wenatchee Basin. The last two sections deal with how the program

---

[1] An efficient monitoring plan reduces "error" to the maximum extent possible. One can think of error as unexplained variability (see Section 3.3), which can reduce monitoring efficiency through the use of invalid statistical designs, biased sampling designs, poorly selected indicators, biased measurement protocols, and non-standardized reporting methods.

will be implemented. Section 8 provides a checklist of questions that need to be addressed in order to implement a valid plan. Section 9 begins to lay out a monitoring plan for the Wenatchee Basin by answering the questions identified in Section 8.

As much as possible, I attempted to keep discussions fairly general. Because this report discusses some issues that are quite involved, I used footnotes to define technical terms, offer further explanation, offer alternative explanations, or to describe a given topic or thought in more detail. I hope the reader will not be too distracted by the extensive use of footnotes. In some instances, it was necessary to provide considerable detail within the text (e.g., discussion on choosing sample sizes).

SECTION 2: STATISTICAL DESIGN

This document defines "statistical design" as the logical structure of a monitoring study. It does not necessarily mean that all studies require rigorous statistical analysis. Rather, it implies that all studies, regardless of the objectives, must be designed with a logical structure that reduces bias and the likelihood that rival hypotheses are correct.[2] My purpose in this section is two-fold. First, I identify the minimum requirements of valid statistical designs and second I identify the appropriate designs for status/trend and effectiveness monitoring. The following discussions draw heavily on the work of Hairston (1989), Hicks et al. (1999), Krebs (1999), Manly (1992, 2001), and Hillman and Giorgi (2002).

Throughout this document I talk about the "validity" of monitoring designs. The validity of a monitoring design is influenced by the degree to which the investigator can exercise experimental control; that is, the extent to which rival variables or hypotheses can be controlled or dismissed. Experimental control is associated with randomization, manipulation of independent variables, sensitivity of dependent (indicator) variables to management activities (treatments), and sensitivity of instruments or observations to measure changes in indicator variables. There are two criteria for evaluating the validity of any effectiveness research design: (1) does the study infer a cause-and-effect relationship (*internal validity*) and (2) to what extent can the results of the study be generalized to other populations or settings (*external validity*)? Ideally, when assessing cause-and-effect, the investigator should select a design strong in both internal and external validity. With some thought, one can see that it becomes difficult to design a study with both high internal and external validity.[3] Because the intent of effectiveness research is to demonstrate a treatment effect, the study should err on the side of internal validity. Without internal validity the data are difficult to interpret because of the confounding effects of uncontrolled variables. Below I identify some common threats to validity.

- Sampling units that change naturally over time, but independently of the treatment, can reduce validity. For example, fine sediments within spawning gravels may decrease naturally over time independent of the treatment. Alternatively, changes in land-use activities upstream from the study area and unknown to the investigator may cause levels of fine sediments to change independent of the treatment.
- The use of unreliable or inconsistent sampling methods or measuring instruments can reduce validity. That is, an apparent change in an indicator variable may actually be nothing more than using an instrument that was not properly calibrated. Changes in indicator variables may also occur if the measuring instrument changes or disturbs the sampling site (e.g., core sampling).
- Measuring instruments that change the sampling unit before the treatment is applied can reduce validity. That is, if the collection of baseline data alters the site in such a

---

[2] Rival hypotheses are alternative explanations for the outcome of an experimental study. In effect, rival hypotheses state that observed changes are due to something other than the management action under investigation.

[3] Studies with high internal validity (laboratory studies) tend to have low external validity. In the same way, studies with high external validity (field studies) tend to have lower internal validity.

way that the measured treatment effect is not what it would be in the population, the results of the study cannot be generalized to the population.

- Differential selection of sampling units can reduce validity, especially if treatment and control sites are substantially different before the study begins. This initial difference may at least partially explain differences after treatment.
- Biased selection of treatment sites can reduce validity. The error here is that the investigator selects sites to be treated in such a way that the treatment effects are likely to be higher or lower than for other units in the population. This issue is complicated by the fact that treatment areas are often selected precisely because they are thought to be problematic.
- Loss of sampling units during the study can reduce validity. This is most likely to occur when the investigator drops sites that shared characteristics such that their absence has a significant effect on the results.
- Multiple treatment effects can reduce validity. This occurs when sampling units get more than one treatment, or the effects of an earlier treatment are present when a later treatment is applied. Multiple treatment effects make it very difficult to identify the treatment primarily responsible for causing a response in the indicator variables.
- The threats above could interact or work in concert to reduce validity.

In most cases, there are simple design elements or requirements that reduce threats to internal and external validity. What follows is a brief description of those elements.

*2.1 **Minimum Requirements***

What are the required elements of a "valid" monitoring study? In general, the more complex the study, the more complex the requirements, but the minimum requirements include *randomization*, *replication*, *independence*, and *controls*.

> **Randomization**—Randomization should be used whenever there is an arbitrary choice to be made of which units will be measured in the sampling frame, or of the units to which treatments will be assigned. The intent is that randomization will remove or reduce systematic errors (bias) of which the investigator has no knowledge. If randomization is not used, then there is the possibility of some unseen bias in selection or allocation. In some situations, complete randomization (both random selection of sampling units and random assignment of treatments) is not possible. Indeed, there will be instances where the investigator cannot randomly assign management activities to survey areas (e.g., removal of mine contaminants from a stream). In this case replication in time and space is needed to generalize inferences of cause-effect relationships.[4] Here, confidence in the inference comes from replication outside the given study area. The rule of thumb is simple: randomize whenever possible.

---

[4] This does not mean that one cannot infer a cause-effect relationship in the study area. The point here is that without random assignment of management activities, it is questionable if results can be generalized to other sites outside the study area.

Replication—Replication is needed to estimate "experimental error," which is the basic unit of measurement for assessing statistical significance or for determining confidence limits. Replication is the means by which natural variability is accounted for in interpreting results. The only way to assess variability is to have more than one replicate for each treatment, including the controls (see Section 3). In the absence of replication, there is no way, without appealing to non-statistical arguments, to assess the importance of observed differences among experimental units. Depending on the objectives of the study, spatial and/or temporal replication may be necessary.

Independence—It is important that the investigator select replicates that are spatially and temporally independent. A lack of independence can confound the study and lead to "pseudoreplication" (Hurlbert 1984). The basic statistical problem of pseudoreplication is that replicates are not independent, and the first assumption of statistical inference is violated. The simplest and most common type of pseudoreplication occurs when the investigator only selects one replicate per treatment. It can be argued that case studies, where a single stream or watershed has been monitored for several years, suffer from pseudoreplication. Therefore, one might conclude that no inference is possible. However, the motive behind a single-replicate case study is different from that behind statistical inference. The primary purpose of a case study is to reveal information about biological or physical processes in the system. This information can then be used to formulate and test hypotheses using real statistical replicates. Indeed, case studies provide the background information necessary to identify appropriate management actions and to monitor their effectiveness.

Investigators need to be aware of spatial pseudoreplication and how to prevent it or deal with it. Spatial pseudoreplication can occur when sampling units are spaced close together. Sampling units close together are likely to be more similar than those spaced farther apart.[5] Spatially dependent sites are "subsamples" rather than replicates and should not be treated as independent replicates. Confounding also occurs when control sites are not independent of treatment sites. This is most likely to occur when control sites are placed downstream from treatments sites (although the reverse can also occur; see Underwood 1994). Understandably, there can be no detection of a management action if the treatment affects both the test and control sites similarly.

Similar, although less often recognized problems occur with temporal replication. In many monitoring studies it is common for sampling to be done once at each of several years or seasons. Any differences among samples may then be attributed to differences among years or seasons. This could be an incorrect inference because a single sample collected each year or season does not account for within year or season variability. Take for example the monitoring of fine sediments in spawning gravels in, say, the Chiwawa River. An investigator measures fine sediments at five random locations (spatial replication) during six consecutive

---

[5] A common concern of selecting sampling units randomly is that there is a chance that some sampling units will be placed next to each other and therefore will lack independence. Although this is true, if the investigator has designed the study so that it accounts for the obvious sources of variation, then randomization is always worthwhile as a safeguard against the effects of unknown factors.

years during the second week of July. A simple statistical analysis of the data could indicate that mean percentages of fine sediments decreased significantly during the latter three years. The investigator may then conclude that fines differed among years.

The conclusion may be incorrect because the study lacked adequate temporal replication. Had the investigator taken samples several times during each year (thereby accounting for within year variability), the investigator may have found no difference among years. A possible reason for the low values during the last three years is because the investigator collected samples before the stream had reached baseflow (i.e., there was a delay in the time that the stream reached baseflow during the last three years compared to the first three years). The higher flows during the second week of July in the last three years prevented the deposition of fines in spawning gravels. An alternative to collecting several samples within years or seasons is to collect the annual sample during a period when possible confounding factors are the same among years. In this case, the investigator could have collected the sample each year during baseflow. The results, however, would apply only to baseflow conditions.

The use of some instruments to monitor physical/environmental indicators may actually lead to pseudoreplication in monitoring designs. This can occur when a "destructive" sampling method is used to sample the same site repeatedly. To demonstrate this point one can look at fine-sediment samples collected repeatedly within the same year. In this example, the investigator designs a study to sample five, randomly-selected locations once every month from June through November (high flows or icing preclude sampling during other months). The investigator randomly selects the week in June to begin sampling, and then samples every fourth week thereafter (systematic sampling). To avoid systematic bias, the same well-trained worker using the same equipment (McNeal core sampler) collects all samples. After compiling and analyzing the data, the investigator may find that there is no significant difference in percent fines among replicates within the year. This conclusion is tenuous because the sampling method (core sampler) disturbed the five sampling locations, possibly reducing fines that would have been measured in following surveys. A more appropriate method would have been to randomly select five new sites (without replacement) during each survey period.

Although replication is an important component of monitoring and should be included whenever possible, it is also important to understand that using a single observation per treatment, or replicates that are not independent, is not necessarily wrong. Indeed, it may be unavoidable in some field studies. What is wrong is to ignore this in the analysis of the data. There are several analyses that can be used to analyze data that are spatially or temporally dependent (see Manly 2001). Because it is often difficult to distinguish between true statistical replicates and subsamples, even with clearly defined objectives, investigators should consult with a professional statistician during the development of monitoring studies.

Controls—Controls are a necessary component of effectiveness research because they provide observations under normal conditions without the effects of the management action

> or treatment. Thus, controls provide the standard by which the results are compared.[6] The exact nature of the controls will depend on the hypothesis being tested. For example, if an investigator wishes to implement a rest-rotation grazing strategy along a stream with heavy grazing impacts, the investigator would monitor the appropriate physical/environmental indicators in both treatment (modified grazing strategy) and control (unmodified intensive grazing) sites. Because stream systems are quite variable, the study should use "contemporaneous controls." That is, both control and treatment sites should be measured at the same time.
>
> Temporal controls can be used to increase the "power" of the statistical design. In this case the treatment sites would be measured before and after the treatment is applied. Thus, the treatment sites serve as their own controls. However, unless there are also contemporaneous controls, all before-after comparisons must assume homogeneity over time, a dubious assumption that is invalid in most ecological studies (Green 1979). Examples where this assumption *is* valid include activities that improve fish passage at irrigation diversions or screen intake structures. These activities do not require contemporaneous controls. However, a temporal control is needed to describe the initial conditions. Therefore, a before-after comparison is appropriate. The important point is that if a control is not present, it is impossible to conclude anything definite about the effectiveness of the treatment.

It should be clear that the minimum requirements of valid monitoring include randomization, replication, independence, and controls. In some instances monitoring studies may lack one or more of these ingredients. Such studies are sometimes called "quasi-experiments." Although these studies are often used in environmental science, they have inherent problems that need to be considered during data analysis. There is no space here to discuss these problems; however, many of them are fairly obvious. The reader should consult Cook and Campbell (1979) for a detailed discussion of quasi-experimental studies.

2.2 Recommended Statistical Designs

A perfect study design would take into account all sources of variability associated with fluctuations in indicator variables. In the absence of perfection, the best approach is to use a design that accounts for all known sources of variation not directly associated with treatment (management action) differences. A reasonable rule is to use the simplest design that provides adequate control of variability. The design should also provide the desired level of precision with the smallest expenditure

---

[6] Lee (1993, pg 205) offers a quote that adequately describes the importance of controls in study designs. Lee writes, "One day when I was a junior medical student, a very important Boston surgeon visited the school and delivered a great treatise on a large number of patients who had undergone successful operations for vascular reconstruction. At the end of the lecture, a young student at the back of the room timidly asked, 'Do you have any controls?' Well, the great surgeon drew himself up to his full height, hit the desk, and said, 'Do you mean did I not operate on half of the patients?' The hall grew very quiet then. The voice at the back of the room very hesitantly replied, 'Yes, that's what I had in mind.' Then the visitor's fist really came down as he thundered, 'Of course not. That would have doomed half of them to their death.' God, it was quiet then, and one could scarcely hear the small voice ask, 'Which half?' (Tufte 1974, p.4--attributed to Dr. E. Peacock, Jr., chairman of surgery, University of Arizona College of Medicine, in Medical World News, Sept. 1, 1974, p. 45.)"

of time and effort. A more complex design has little merit if it does not improve the performance of statistical tests or provide more precise parameter estimates. Furthermore, an efficient design usually leads to simpler data analysis and cleaner inferences. Below I describe valid designs for both effectiveness and status/trend monitoring.

> **Effectiveness Monitoring**—Because effectiveness monitoring attempts to explain cause-and-effect relationships (e.g., effect of a tributary project on fish abundance), it is important to include as many of the elements of valid statistical design as possible. An appropriate design recommended by the Action Agencies/NOAA Fisheries (2003), ISAB (2003), and WSRFB (2003) is the Before-After-Control-Impact or BACI design (Stewart-Oaten et al. 1986, 1992; Smith et al. 1993). This type of design is also known as a Control-Treatment Paired or CTP design (Skalski and Robson 1992), or Comparative Interrupted Time Series design (Manly 1992). Although names differ, the designs are essentially the same. That is, they require data collected simultaneously at both treatment and control sites before and after treatment. These data are paired in the sense that the treatment and control sites are as similar as possible and sampled simultaneously. Replication comes from collecting such paired samples at a number of times (dates) both before and after treatment. Spatial replication is possible if the investigator selects more than one treatment and control site.[7] The pretreatment sampling serves to evaluate success of the pairings and establishes the relationship between treatment and control sites before treatment. This relationship is later compared to that observed after treatment.
>
> The success of the design depends on indicator variables at treatment and control sites "tracking" each other; that is, maintaining a constant proportionality. The design does not require exact pairing; indicators simply need to "track" each other. Such synchrony is likely to occur if similar climatic and environmental conditions equally influence sampling units. Precision of the design can be improved further if treatment and control stream reaches are paired according to a hierarchical classification approach (see Section 4). Thus, indicator variables in stream reaches with similar climate, geology, geomorphology, and channel types should track each other more closely than those in reaches with only similar climates.
>
> It is important that control and treatment sites be independent; treatment at one site cannot affect indicators in another site. The NRC (1992) recommends that control data come from another stream or from an independent reach in the same stream. After the pretreatment period, sites to be treated should be selected randomly.[8] Randomization eliminates site location as a confounding factor and removes the need to make model-dependent inferences (Skalski and Robson 1992). Hence, conclusions carry the authority of a "true" experiment and will generally be more reliable and less controversial. Post-treatment observations should be made simultaneously in both treatment and control sites.

---

[7] The use of several test and control sites is recommended because it reduces spatial confounding. In some instances it may not be possible to replicate treatments, but the investigator should attempt to replicate control sites. These "Beyond BACI" designs and their analyses are described in more detail in Underwood (1996).

[8] As noted later, in most cases treatments will not be randomly assigned to sites. Thus, the studies will be "causal-comparative," rather than "true" experimental studies.

Several different statistical procedures can be used to analyze BACI designs. Manly (1992) identified three methods: (1) a graphical analysis that attempts to allow subjectively for any dependence among successive observations, (2) regression analysis, which assumes that the dependence among successive observations in the regression residuals is small enough to ignore, and (3) an analysis based on a time series model that accounts for dependence among observations. Cook and Campbell (1979) recommend using autoregressive integrated moving average models and the associated techniques developed by Box and Jenkins (1976). Skalski and Robson (1992) introduced the odd's-ratio test, which looks for a significant change in dependent variable proportions in control-treatment sites between pretreatment and post-treatment phases. A common approach, recommended by WSRFB (2003), includes analysis of difference scores. Differences are calculated between paired control and treatment sites. These differences are then analyzed for a before-after treatment effect with a two-sample t-test, Welch modification of the t-test, or with nonparametric tests like the randomization test, Wilcoxon rank sum test, or the Mann-Whitney test (Stewart-Oaten et al. 1992; Smith et al. 1993). Choice of test depends on the type of data collected and whether those data meet the assumptions of the tests.

In some cases, the investigator will not be able to randomly assign treatments to sampling locations. Despite a lack of randomization of treatment conditions, if the treatment conditions are replicated spatially or temporally, a sound inference to effects may be possible. Although valid statistical inferences can be drawn to the sites or units, the authority of a randomized design is not there to "prove" cause-effect relationships. Skalski and Robson (1992) describe in detail how to handle BACI designs that lack randomization.

Status/Trend Monitoring—Because the intent of status/trend monitoring is simply to describe existing conditions and document changes in conditions over time, it does not require all the elements of valid statistical design found in effectiveness monitoring studies. For example, controls are not required in status/trend monitoring. Controls would be important if one desires to assess cause-and-effect relationships (goal of effectiveness monitoring), which is not the purpose of status/trend monitoring. However, status/trend monitoring does require temporal and spatial replication and probabilistic sampling.

Monitoring the status and trends of Evolutionarily Significant Units (ESUs), populations, subpopulations, and habitat characteristics is an important component of the Action Agencies/NOAA Fisheries RME Plan, which will be implemented within the Wenatchee Basin. The Plan calls for the implementation of the U.S. Environmental Protection Agency's Environmental Monitoring and Assessment Program (EMAP) design, which is a spatially-balanced, site-selection process developed for aquatic systems. The state of Oregon has successfully implemented an EMAP-based program for coastal coho salmon (Moore 2002). The monitoring program as implemented in Oregon is spatially explicit, unbiased, and has reasonably high power for detecting trends. The design is sufficiently flexible to use on the scale of multiple large river basins and can be used to estimate the numbers of adult salmon returning each year, the distribution and rearing density of juvenile salmon, productivity and relative condition of stream biota, and freshwater habitat conditions. In addition, the EMAP site-selection approach supports sampling at varying spatial extents.

Specifically, EMAP is a survey design that was developed to describe current status and to detect trends in a suite of indicators. These two objectives have conflicting design criteria; status is ordinarily best assessed by including as many sample units as possible, while trend is best detected by repeatedly observing the same units over time (Overton, et al. 1990). EMAP addresses this conflict by using rotating panels (Stevens 2002). Each panel consists of a collection of sites that will have the same revisit schedule over time. For example, sites in one panel could be visited every year, sites in another revisited every five years, and sites in still another revisited every ten years. As a starting point for the Wenatchee Basin, it is recommended that the design include six panels, with one panel defining sites visited every year and five panels defining sites visited on a five-year cycle (Table 1). The process by which sites are selected for each panel and the statistical methods used to analyze data are described in Section 3.

Table 1. Shading indicates the years in which sites within each panel are sampled in the Wenatchee Basin. For example, sites in panel 1 are visited every year, while sites in panel 2 are visited only in years 1, 6, 11, and 16, assuming a 20-year sampling frame.

	Year																			
Panel	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
1	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨
2	▨					▨					▨					▨				
3		▨					▨					▨					▨			
4			▨					▨					▨					▨		
5				▨					▨					▨					▨	
6					▨					▨					▨					▨

SECTION 3: SAMPLING DESIGN

Once the investigator has selected a valid statistical design, the next step is to select "sampling" sites. *Sampling* is a process of selecting a number of units for a study in such a way that the units represent the larger group from which they were selected. The units selected comprise a *sample* and the larger group is referred to as a *population*.[9] All the possible sampling units available within the area (population) constitute the *sampling frame*.[10] The purpose of sampling is to gain information about a population. If the sample is well selected, results based on the sample can be generalized to the population. Statistical theory assists in the process of drawing conclusions about the population using information from a sample of units.

Defining the population and the sample units may not always be straightforward, because the extent of the population may be unknown, and natural sample units may not exist. For example, a researcher may exclude livestock grazing from sensitive riparian areas in a watershed where grazing impacts are widespread. In this case the management action may affect aquatic habitat conditions well downstream from the area of grazing. Thus, the extent of the area (population) that might be affected by the management action may be unclear, and it may not be obvious which sections of streams to use as sampling units.

When the population and/or sample units cannot be defined unambiguously, the investigator must subjectively choose the potentially affected area and impose some type of sampling structure. For example, sampling units could be stream habitat types (e.g., pools, riffles, or glides), fixed lengths of stream (e.g., 100-m long stream reaches), or reach lengths that vary according to stream widths (e.g., see Simonson et al. 1994). Before selecting a sampling method, the investigator must define the population, size and number of sample units, and the sampling frame.

3.1 Methods of Selecting a Sample

Selection of a sample is a crucial step in monitoring fish populations and physical/environmental conditions in streams. The "goodness" of the sample determines the generalizability of the results. Because monitoring studies usually require a large amount of time and money, non-representative results are wasteful. Therefore, it is important to select a method or combination of methods that increases the degree to which the selected sample represents the population. Below I describe the five most commonly used methods for monitoring fish populations and physical/environmental conditions: random sampling, stratified sampling, systematic sampling, cluster sampling, and multi-stage sampling. See Scheaffer et al. (1990) for a more detailed discussion of sampling methods.

---

[9] This definition makes it clear that a "*population*" is not limited to a group of organisms. In statistics, it is the total set of elements or units that are the target of our curiosity.

[10] The *sampling frame* is a "list" of all the available units or elements from which the sample can be selected. The sampling frame should have the property that every unit or element in the list has some chance of being selected in the sample. A sampling frame does not have to list all units or elements in the population.

Random sampling—A simple random sample is one that is obtained in such a way that all units in the defined sampling frame have an equal and independent chance of being selected. Stated differently, every unit has the same probability of being selected and the selection of one unit in no way affects the selection of another unit. Random sampling is the best single way to obtain a representative sample.[11] Random sampling should lead to small and unsystematic differences between the sample and the population because differences are a function of chance and not the result of any conscious or unconscious bias on the part of the investigator. Random sampling is also required by inferential statistics. This is important because statistics permit the researcher to make inferences about populations based on the behavior of samples. If samples are not randomly selected, then one of the major assumptions of inferential statistics is violated, and inferences are correspondingly tenuous.

The process of selecting a random sample involves defining the sampling frame, identifying each unit within the frame, and selecting units for the sample on a completely chance basis. If the sampling frame contains units numbered from 1 to **N**, then a simple random sample of size **n** is obtained without replacement by drawing **n** numbers one by one in such a way the each choice is equally likely.

Stratified sampling—Stratified sampling is the process of selecting a sample in such a way that identified strata in the sampling frame are represented in the sample.[12] This sampling method addresses the criticism that simple random sampling leaves too much to chance, so that the number of sampling units in different parts of the population may not match the distribution in the population.

Stratified sampling involves dividing the units in the sampling frame into non-overlapping strata, and selecting an independent random sample from each of the strata. An example would be to stratify a stream based on habitat types (i.e., pools, riffles, glides, etc.) and then randomly select **n** units within each habitat type. This would ensure that each habitat type is represented in the sample. There are a couple of advantages of stratified sampling: (1) if the sampling units within the strata are more similar than units in general, the estimate of the overall population mean will have a smaller standard error than a mean calculated with simple random sampling; and (2) there may be value in having separate estimates of population parameters for the different strata. Stratification requires the investigator to consider spatial location, areas within which the population is expected to be uniform, and the size of sampling units. Generally, the choice of how to stratify is just a question of common sense.

In some situations there may be value in analyzing a simple random sample as if it were obtained by stratified random sampling. That is, one takes a simple random sample and then places the units into strata, possibly based on information gathered at the time of sampling.

---

[11] No sampling technique guarantees a representative sample, but the probability is higher for random sampling than for other methods.

[12] The number of units selected from each strata could be equal (i.e., n is the same for all strata), or the number could be proportional to the size of the strata. Equal-sized samples would be desired if one wanted to compare the performance of different strata.

The investigator then analyzes the sample as if it were a stratified random sample. This procedure is known as *post-stratification*. Because a simple random sample should place sample units in different strata according to the size of those strata, post-stratification should be similar to stratified sampling with proportional allocation, provided the total sample size is reasonably large. This may be valuable particularly when the data may be used for a variety of purposes, some of which are unknown at the time of sampling.

Systematic sampling—Systematic sampling is sampling in which units are selected from a list by taking every k^{th} unit. If k = 4, one would sample every 4th unit; if k = 10, one would sample every 10th unit. The value of **k** depends on the size of the sampling frame (i.e., the total number of units) and the desired sample size. The major difference between systematic sampling and the methods discussed above is that all units of the population do not have an independent chance of being selected. Once the first unit is selected, all remaining units to be included in the sample are automatically determined. Nevertheless, systematic sampling is often used as an alternative to simple random sampling or stratified sampling for two reasons. First, the process of selecting sample units is simpler for systematic sampling. Second, under certain circumstances, estimates for systematic sampling may be more precise because the population is covered more evenly. Systematic sampling is not recommended if the population being sampled has some cyclic variation (e.g., regular occurrence of pools and riffles along the course of a stream). Simple random sampling and stratified sampling are not affected by patterns in the population.

Cluster sampling—Cluster sampling is sampling in which groups, not individual units, are randomly selected. Thus, cluster sampling involves sampling clusters of units rather than single units. All units of selected groups have similar characteristics. For example, instead of randomly selecting pools throughout a watershed, one could randomly select channel bed-form types (e.g., plane-bed, step-pool, etc.) within the watershed and use all the pools within those randomly-selected channel types. Cluster sampling is more convenient when the population is very large or spread out over a wide geographic area. This advantage is offset to some extent by the tendency of sample units that are close together to have similar measurements. Therefore, in general, a cluster sample of **n** units will give estimates that are less precise than a simple random sample of **n** units. Cluster sampling can be combined with stratified sampling (see Scheaffer et al. 1990 for more details).

Multi-Stage Sampling—Multi-stage sampling is sampling in which clusters or stages (and clusters within clusters) are randomly selected and then sample units are randomly selected from each sampled cluster. With this type of sampling, one regards sample units as falling within a hierarchical structure. The investigator randomly samples at each of the various levels within the structure. For example, suppose that an investigator is interested in describing changes in fine sediments in stream riffles after livestock grazing is removed from sensitive riparian areas in a large watershed. The investigator may be able to divide the watershed into different geological/geomorphic units (primary sampling units) and then divide each geological/geomorphic unit into channel types (secondary sampling unit). Finally, the investigator may divide each channel type into habitat types (e.g., pools, riffles, glides, etc.). The investigator would obtain a "three-stage" sample of riffle habitats by first randomly selecting several primary sampling units (geological/geomorphic units), next randomly selecting one or more channel types (second-stage units) within each sampled primary unit, and finally randomly selecting one or more riffles (third-stage units) from each sampled channel type. This type of sampling is useful when a hierarchic structure exists, or when it is simply convenient to sample at two or more levels.

It is important to note that some monitoring programs include a combination of sampling designs. As you will see later in this section, the EMAP approach is a combination of random and systematic sampling. Juvenile fish monitoring in the Chiwawa Basin included a combination of stratified random sampling and two-stage sampling (Hillman and Miller 2002). These complex sampling designs require an understanding of the more basic designs.

3.2 Choosing Sample Size

I now address the question, "to have a high probability of detecting a management (treatment) effect (effectiveness monitoring) or a change in current conditions (status/trend monitoring), what sample size should the investigator use?" This is one of the most important questions of a monitoring plan. If the sample is too small, the results of the study may not be generalizable to the population. In addition, the wrong decision may be made concerning the validity of the hypothesis. Therefore, it is important that the investigator select a sample size that will increase the validity of the hypothesis. Fortunately, there are a number of equations and tables that can assist in selecting sample sizes. Before I consider these, it is appropriate to discuss the factors that one needs to consider when selecting a total sample size.

In general, the total sample size for status/trend monitoring depends upon the population size (total number of units in the sampling frame), population variance or standard deviation, and the level of error that the investigator considers acceptable. Quite often the population standard deviation is unknown. In this situation, the investigator can replace the population standard deviation with the sample standard deviation, which may be available from previous studies (an informal "meta-analysis"). Scheaffer et al. (1990) and Browne (2001) describe methods for guessing the population standard deviation when little prior information is available.[13] The level of error is selected by the investigator and should be based on the objectives of the study. Many studies set the error at 0.05. Scheaffer et al. (1990) provide equations for estimating sample sizes for simple random, stratified, systematic, and cluster sampling. There are also a number of computer packages that can be used to estimate sample sizes, such as PASS 2000 (Power Analysis and Sample Size), which is produced by NCSS Statistical Software (2000), and Methodologist's Toolchest, which is produced by Idea Works (1997).[14]

---

[13] For simple random sampling, the guess is one-fourth the range of possible values. The idea being that for many distributions the effective range is the mean plus and minus about two standard deviations. This type of approximation is often sufficient because it is only necessary to get the sample size roughly right.

[14] The use of trade or firm names in this paper is for reader information only and does not imply endorsement by an agency of any product or service.

Effectiveness monitoring, on the other hand, almost always requires the testing of statistical hypotheses, which means that additional factors must be considered when selecting a total sample size. Indeed, statistical significance is usually the desired outcome of effectiveness monitoring (i.e., statistical significance indicates that the management action did what it was suppose to do).[15] Therefore, when selecting a total sample size for effectiveness monitoring, the investigator must carefully evaluate all the factors that influence the validity of statistical hypotheses. These factors include significance level, effect size, variability, and statistical power.[16] What follows is a brief description of each of these factors. First, however, I briefly describe the errors of inference.

> **Errors of Inference**—There are four possible outcomes of a statistical hypothesis test. If the hypothesis of no difference (null hypothesis) is really true, then two outcomes are possible: not rejecting the null hypothesis is a correct inference, while rejecting it constitutes a Type I error. That is, a Type I error occurs when the investigator concludes that a difference between or among treatments is real when in fact it is not. Similarly, if the null hypothesis is really false, the correct inference is to reject it, and failing to do so constitutes a Type II error. To quickly recap, a Type I error occurs when the investigator concludes that a difference is real when in fact it is not. A Type II error occurs when the investigator concludes that there is no difference when in fact a difference exists. In statistical terms, the probability of committing a Type I error is α, while the probability of a Type II error is β. The power of the test (**1-β**) is the probability of correctly rejecting the null hypothesis when it is really false.
>
> Both types of errors can be costly in monitoring studies where management actions involve the effects of commercial activities, such as timber harvesting or road building, on stream ecosystems. For example, a Type I error may lead to unnecessary limitations on commercial activities, while a Type II error may result in the continuation of activities damaging to the stream ecosystem. While it is impossible to calculate the probability that a hypothesis is true using classical statistical tests, the probability of incurring either a Type I or a Type II error can be controlled to acceptable levels. For example, Type I error is typically limited by the conventional significance level of statistical tests to a frequency of less than five errors per 100 tests performed (“**critical α**”<0.05). In other words, a critical α of 0.05 means that if the null hypothesis was really true and the experiment was repeated many times, the null hypothesis would be rejected incorrectly in at most 5% of the replicate experiments. In contrast, “statistical power analysis” is used to estimate and limit Type II error.
>
> **Significance Level**—The significance level is a critical value of α, which is the maximum probability of a Type I error that the researcher is willing to accept. When a P-value is less than 0.05 (the usual critical value of α), the researcher rejects the null hypothesis with the guarantee that the chance is less than 1 in 20 that a true null hypothesis has been rejected. Of course, this guarantee about the probability of making a Type I error is valid only if the

---

[15] As I pointed out earlier, not all effectiveness research requires the testing of statistical hypotheses. For example, improving fish passage at a culvert or irrigation diversion does not require one to test a statistical hypothesis. It does require that the results of the action comply with the desired outcome.

[16] Total sample size is also affected by the choice of experimental design and statistical analysis. Because these two factors are used to explain or partition variability, I included them in my discussion on variability and in Section 2.

assumptions of the test are met. The probability of a Type I error (significance level) is completely under the control of the investigator and is inversely related to total sample size. However, increasing critical α-level is not the most effective way to reduce total sample size or to gain statistical power (Lipsey 1990). Generally one increases the significance level when the cost of Type II errors is much larger than the cost of Type I errors.

Effect size—The effect size is the size of change in the parameter of interest that can be detected by an experiment. In statistical jargon, effect size is the difference between the equality components of the null and alternative hypotheses, usually chosen to represent a biologically or practically significant difference.[17] For example, a practical significant effect size of interest might be the difference between the maximum acceptable percentage of fine sediments in spawning gravels and the current percentage of fines in spawning gravels. The investigator must select an effect size to calculate total sample size.

Selection of significant effect size can be straightforward for some designs. In the example above, the practical significant effect size was the difference between a population mean and a known constant (e.g., maximum acceptable percentage of fines in spawning gravels). Similarly, when comparing two population means or two correlation coefficients, the estimate of effect size is simply the difference between the two values. However, formulas for effect size become more complex in designs that involve many relationships among statistical parameters, such as analysis of variance or multiple regression.

In other cases the selection of an appropriate effect size is difficult because it is very subjective. Ideally the effect size to be detected should be practically significant, but quite often this value cannot be expressed quantitatively because of a lack of information. In the absence of information, Cohen (1988) proposes small, medium, and large standardized effect sizes. Standardized effect sizes include measures of variance as well as summaries of the magnitude of treatment effects. For example, the standardized effect size for the difference between two means is expressed as the effect size (μ_1-μ_2), divided by the common standard deviation (σ). According to Cohen (1988), small effects sizes [$(\mu_1-\mu_2)/\sigma=0.2$] are subtle, medium effect sizes (0.5) are large enough to be perceived in the course of normal experience, and large effect sizes (0.8) are easily perceived at a glance. One should use caution when selecting standardized effect sizes based on Cohen. His standardized effect sizes are derived from behavioral studies, which may not represent ecological studies. In general, sample size is inversely related to effect size. In other words, a larger sample size is needed to detect a small significant effect size.

Variability—Variability is a measure of how much scores (e.g., water temperatures) differ (vary) from one another. A measure of variability simply indicates the degree of dispersion among the set of scores. If the scores are similar, there is little dispersion and little variability.

---

[17] Often, statistical significance and biological significance differ. For example, a temperature difference of 0.2ºC may be significant statistically, but not biologically. On the other hand, a 1.0ºC may be biologically significant, but because of a small sample size, the difference is not significant statistically. It is important that the investigator design the study to assess biological or practical significance.

If the scores are dissimilar, there is a high degree of dispersion (variability). In short, a measure of variability does nothing more than indicate the spread of scores. The variance and the standard deviation are often used to describe the variability among a group of scores. An estimate of the population variability is generally needed to calculate sample size. As we indicated earlier, if the population standard deviation is not available, one can use the sample standard deviation (from other studies or pilot studies) as an estimate of the population standard deviation, or one can guess the variability using methods described in Scheaffer et al. (1990).[18] In general, the greater the variability the larger the sample size needed to detect a significant difference.

Statistical Power—Statistical power is the probability that a statistical test will result in statistical significance (Cohen 1988). More technically, statistical power (**1-β**) is the probability of detecting a specified treatment effect (management action) when it is present. Its complement, β, is the probability of a Type II error. Sample size is directly proportional to statistical power. That is, greater statistical power requires a larger sample size. Cohen (1988) suggested that experiments should be designed to have a power of 0.80 ($\beta = 0.20$). This comports with Peterman (1990) and Green (1994), who suggest that fisheries researchers should prefer β at least <0.2, or power ≥ 0.8. If the investigator desires to be as conservative about making Type II as Type I errors, β should equal α, or desired power = 0.95 if $\alpha = 0.05$ (Lipsey 1990).

In summary, significance level, effect size, variability, and statistical power affect the total sample size needed for most effectiveness monitoring studies. Because of the time and cost of sampling fish and physical/environmental conditions in tributary habitats, it should be the desire of the investigator to sample the minimum possible number of units. There are several ways that one can reduce sample size. One can reduce statistical power, increase effect size, decrease the variance of the observed variables, or increase the probability of making a Type I error. Although any one of these can be used to reduce the total sample size, it is not necessarily wise (or even possible) to manipulate all of them.

Alpha is completely under the control of the researcher and there may be good reasons to choose critical α-levels other than 0.05. However, changing the critical α-level is not the most effective way to reduce sample size (Lipsey 1990). In addition, it is unwise to reduce statistical power (1-β), unless there is good reason to do so. The objective of the study should guide the value of α and β. Data snooping or exploratory research, for example, will often be more cost-effective if α is set relatively high and β relatively low, because the objective is to detect previously unknown relationships. In addition, one should consider the prior probability that each hypothesis is true. A hypothesis that seems likely to be true, based on previous work, should be treated more cautiously with respect to

---

[18] If there are no estimates of variability, one can use the "signal-to-noise ratio" to estimate sample size (see Green 1994). The signal-to-noise ration is the ratio of the effect size to standard deviation. This approach may be appealing because an estimate of population variability seems to disappear, as does the need to estimate it. However, I do not recommend using this ratio to calculate sample size because it really does matter what the standard deviation is. The standard deviation is partly natural variation, but it also contains sampling and analysis error. The latter sources of error will affect the estimate of total sample size. Furthermore, to some degree the investigator can control the size of the standard deviation (by using valid designs and selecting sensitive indicators and reliable measurements). Therefore it is best to have some estimate of population standard deviation.

erroneous rejection than a hypothesis that seems less credible (Lipsey 1990). Mapstone (1995) offers a method of selecting α and β based on the relative weighting of the perceived consequences of Type I and Type II errors. I recommend that investigators review the methods proposed in Mapstone (1995).

Increasing effect size and/or decreasing variability may be the most effective ways to reduce sample size. However, the investigator has little flexibility in selecting significant effect sizes. Effect size is based on "practical significance" or the difference between some desirable condition and current conditions. It is inappropriate to "stretch" the effect size beyond what is considered practically significant. Consequently, the investigator is left primarily with reducing variability as a means of reducing sample size. Because physical/environmental variables often exhibit large variances, strategies for reducing variability are especially important for reducing sample size (and achieving high statistical power). Variability is generally reduced by improving measurement precision, selecting dependent (indicator) variables that are sensitive to the management action, and by various techniques of experimental design (e.g., blocking,[19] stratification, or covariate analysis). Later I will identify sensitive indicator variables (Section 4) and reliable methods for measuring those variables (Section 5).

There are a number of aids that the investigator can use to estimate total sample size. Cohen (1988) provides tables and equations for calculating sample sizes. Various computer packages also estimate sample sizes, such as PASS 2000, SYSTAT, and Methodologist's Toolchest. I suggest that the investigator use the method that meets their particular needs.

3.3 Measurement Error

Measurements and estimates are never perfect. Indeed, most fish population and habitat variables are difficult to measure, and the errors in these measurements are often large. It is tempting to ignore these errors and proceed as though the estimates reflect the true state of the resource. One should resist this temptation because it could lead to missing a treatment effect, resulting in a waste of money and effort. Investigators need to be aware of the types of errors and how they can be identified and minimized. This is important because total sample size and statistical power are related to variability. By reducing measurement error and bias, one effectively reduces variability, resulting in greater statistical power. In this section I identify and describe the various types of errors and describe ways to minimize these errors.

In general, "error" indicates the difference between an estimated value (from a sample) and its "true" or "expected" value. The two common types of error are *random error* and *systematic error*. Random error (a.k.a. chance error) refers to variation in a score or result that displays no systematic *bias*[20] when taking repeated samples. In other words, random error is the difference between the

---

[19] Although unreplicated random block designs are useful methods of reducing variability, I do not recommend them for monitoring tributary conditions because they fail to deal with interactions between treatments (management actions) and blocks. The assumption of no interaction is unrealistic in environmental studies (Underwood 1994).

[20] *Bias* is a measure of the divergence of an estimate (statistic) from the population parameter in a particular direction. The greater the divergence the greater the bias. Nonrandom sampling often produces such bias.

estimate of a population parameter that is determined from a random sample and the true population value, absent any systematic bias. One can easily detect the presence of random errors by simply repeating the measurement process several times under similar conditions. Different results, with no apparent pattern to the variation (no bias) indicate random error. Although random errors are not predictable, their properties are understood by statistical theory (i.e., they are subject to the laws of probability and can be estimated statistically). The standard deviation of repeated measurements of the same phenomenon gauges the average size of random errors.[21]

Random errors can occur during the collection and compilation of sample data. These errors may occur because of carelessness in recording field data or because of missing data. Recording errors can occur during the process of transferring information from the equipment to field data sheets. This often results from misplacing decimal points, transposing numbers, mixing up variables, or misinterpreting hand-written records. Although not always the fault of the investigator, missing data are an important source of error.

Systematic errors or bias, on the other hand, are not subject to the laws of probability and cannot be estimated or handled statistically without an independent estimate of the bias. Systematic errors are present when estimates consistently over or underestimate the true population value. An example would be a poorly calibrated thermometer that consistently underestimates the true water temperature. These errors are often introduced as a result of poorly calibrated data-recording instruments, miscoding, misfiling of forms, or some other error-generating process. They may also be introduced via interactions among different variables (e.g., turbidity is usually highest at high flows). Systematic error can be reduced or eliminated through quality control procedures implemented at the time data are collected or through careful checking of data before analysis. For convenience, I divided systematic errors into two general classes: those that occur because of inadequate procedures and those that occur during data processing. I consider each of these in turn.

> **Biased Procedures**—A biased procedure involves problems with the selection of the sample, the estimation of population parameters, the variables being measured, or the general operation of the survey. For example, selecting sample units based on access can increase systematic error because the habitat conditions near access points may not represent the overall conditions of the population. Changing sampling times and sites during the course of a study can introduce systematic error. Systematic errors can grow imperceptibly as equipment ages or observers change their perspectives (especially true of "visual" measurements). Failure to calibrate equipment introduces error, as does demanding more accuracy than can be expected of the instrument or taking measurements outside the range of values for which the instrument was designed.
>
> **Processing Errors**—Systematic errors can occur during compiling and processing data. Errors can occur during the transfer of field records to computer spreadsheets. Investigators

---

[21] It is important not to confuse standard deviation with standard error. The *standard error of a sample average* gauges the average size of the fluctuation of means from sample to sample. The *sample standard deviation* gauges the average size of the fluctuations of the values within a sample. These two quantities provide different information.

can also introduce large systematic errors by using faulty formulas (e.g., formulas for converting variables). Processing errors are the easiest to control.

The investigator must consider all these sources of error and develop a plan (quality control plan) that minimizes measurement bias. Certainly some errors are inevitable, but a substantial reduction in systematic errors will benefit a monitoring study considerably. I offer the following guidelines for achieving this goal.

(1) Measures based on counts (e.g., Redds, LWD, Pools)

- Make sure that new personnel are trained adequately by experienced workers.
- Reduce errors by taking counts during favorable conditions and by implementing a rigorous protocol.
- If an over or underestimate is assumed, attempt to assess its extent by taking counts of populations of known size.

(2) Measures based on visual estimates (e.g., snorkel surveys, bank stability)

- Make sure that all visual estimates are conducted according to rigorous protocols by experienced observers.
- Attempt to assess observer bias by using trained personnel to check observations of new workers.

(3) Measures based on instruments (e.g., dissolved oxygen, temperature)

- Calibrate instruments before first use and periodically thereafter.
- Personnel must be trained in the use of all measuring devices.
- Experienced workers should periodically check measurements taken by new personnel.
- Use the most reliable instruments.

(4) Re-measurement of indicators

- Use modern GPS technology and carefully marked maps and diagrams to relocate previous sampling units.
- Guard against the transfer of errors from previous measurements.
- Make sure that bias is not propagated through the use of previous measurements as guides to subsequent ones.

(5) Handling of data

- Record data directly into electronic form where possible.
- Back-up all data frequently
- Design manual data-recording forms and electronic data-entry interfaces to minimize data-entry errors.

- Use electronic data-screening programs to search for aberrant measurements.
- Frequently double-check the transfer of data from field data forms to computer spreadsheets.

Before I leave this discussion, it is important to describe briefly how one should handle outliers. Outliers are measurements that look aberrant (i.e., they appear to lie outside the range of the rest of the values). Because they stand apart from the others, it appears as if the investigator made some gross measurement error. It is tempting to discard them not only because they appear unreasonable, but because they also draw attention to possible deficiencies in the measurement process. Before discarding an apparent outlier, the investigator should look thoroughly at how they were generated. Quite often apparent outliers result from simple errors in data recording, such as a misplaced decimal point. On the other hand, they may be part of the natural variability of the system and therefore should not be ignored or discarded.[22] If one routinely throws out aberrant values, the resulting data set will give false impressions of the structure of the system. Therefore, as a general rule, investigators should not discard outliers unless it is known for certain that measurement errors attend the estimates.

3.4 Recommended Sampling Designs

Using the basic tools described above, valid sampling designs can be identified for status/trend and effectiveness monitoring in the Wenatchee Basin. The recommended sampling designs, if implemented correctly, should reduce bias and error.

> **Effectiveness Monitoring**—This plan recommends that sampling units for effectiveness monitoring be selected according to a stratified random sampling design. The plan requires that streams or stream segments to be treated with some action(s) will be classified according to a hierarchical classification system (see Section 4). Once classification identifies non-overlapping strata, sampling sites are then selected randomly within each stratum. The same process occurs within control or reference areas, which are similar to treatment areas based on classification. The number of sites selected will depend on effect size, variability, power, and significance levels. The number of sites within each stratum should be proportional to the size of the stratum. That is, a larger stratum would receive more sites than a smaller stratum.
>
> **Status/Trend Monitoring**—Because the plan follows EMAP, which requires spatially balanced samples, sites will be selected according to the generalized random tessellation stratified design (GRTS) (Stevens 1997; Stevens and Olsen 1999; Stevens and Urquhart 2000; Stevens 2002). Briefly, the GRTS design achieves a random, nearly regular sample point pattern via a random function that maps two-dimensional space onto a one-dimensional line (linear space). A systematic sample is selected in the linear space, and the sample points are mapped back into two-dimensional space. The GRTS design is used to select samples for all panels (six panels for the Wenatchee Basin).

---

[22] Another reason that outliers should be treated carefully is because they can invalidate standard statistical inference procedures. Outliers tend to affect assumptions of variability and normality.

As a starting point, the plan recommends a sample size of 25 sites per panel. This means that GRTS will select a total of 150 sites (6 panels x 25 sites per panel = 150 sites). Two panels of sites will be monitored each year (see Section 2.2), resulting in a total of 50 sites sampled annually within the Wenatchee Basin. Some of the 150 sites selected may fall in areas that are physically inaccessible or cannot be accessed because of landowner denial. Therefore, GRTS will select an additional 15 sites, any one of which can replace an inaccessible site.

The sampling frame for the 150 sites will consist of all second through fifth-order streams in the Wenatchee Basin. These stream segments were selected because most spawning and rearing of ESA-listed fish species occur in these areas. Because it is unclear at this time which stream segments (orders) should receive the highest density of sampling sites, a variety of scenarios will be modeled (Table 2). The first is an equal number of sites among the stream orders, the second gives more weight (higher density of sites) to third and fourth order streams, while the last gives the greatest weight to fourth order streams. The results of these scenarios will be evaluated to see which one most closely fits the objectives of status/trend monitoring in the Wenatchee Basin.

Table 2. Proportion of sample sites distributed among stream orders within the Wenatchee Basin.

Scenario	**Stream order**			
	2	**3**	**4**	**5**
1	0.25	0.25	0.25	0.25
2	0.20	0.30	0.30	0.20
3	0.20	0.20	0.40	0.20

Data collected within the EMAP design will be analyzed according to the statistical protocols outlined in Stevens (2002). The Horvitz-Thompson or π-estimator is recommended for estimation of population status. Multi-phase regression analyses are recommended for estimating the distribution of trend statistics. These approaches are fully explained in Diaz-Ramos et al. (1996) and Stevens (2002).

SECTION 4: SAMPLING AT DIFFERENT SPATIAL SCALES

Because monitoring will occur at a range of spatial scales, there may be some confusion between the roles of status/trend monitoring and effectiveness monitoring. Generally, one thinks of status/trend monitoring as monitoring that occurs at coarser scales and effectiveness monitoring at finer scales. In reality, both occur across different spatial scales, and the integration of both is needed to develop a valid monitoring program (ISAB 2003; AA/NOAA Fisheries 2003; WSRFB 2003).

The scale at which status/trend and effectiveness monitoring occurs depends on the objectives of the study, the size or distribution of the target population, and the indicators that will be measured. In status/trend monitoring, for example, the objective may be to measure egg-parr survival of spring chinook salmon. Because the Wenatchee Basin consists of one population of spring chinook, the entire basin is the spatial scale at which egg-parr survival is monitored. In contrast, if the objective is to assess egg-parr survival of spring chinook in the Chiwawa Basin (a sub-population of the Wenatchee population), the spatial scale at which monitoring occurs includes only the Chiwawa Basin, a much smaller area than the entire Wenatchee Basin. Thus, status/trend monitoring can occur at various scales depending on the distribution of the population of interest.

In the same way, effectiveness monitoring can occur at different spatial scales. That is, one can assess the effect of a tributary action on a specific ESU (which may encompass several populations), a specific population (may include several sub-populations), at the sub-population level (may encompass a watershed within a basin), or at the reach scale. Clearly, the objectives and hence the indicators measured dictate the spatial scale at which effectiveness monitoring is conducted. For example, if the objective is to assess the effects of nutrient enhancement on egg-smolt survival of spring chinook in the Chiwawa Basin (a sub-population of the Wenatchee spring chinook population), then the spatial scale covered by the study must include the entire area inhabited by the eggs, fry, parr, and smolts. If, on the other hand, the objective is to assess the effects of a sediment reduction project on egg-fry survival of a local group of spring chinook (i.e., chinook within a specific reach of stream), then the study area would only encompass the reach of stream used by spawners of that local group.

In theory there might be no limit to the scale at which effectiveness monitoring can be applied, but in practice there is a limit. This is because as the spatial scale increases, the tendency for multiple treatments (several habitat actions) affecting the same population increases (Table 3). That is, at the spatial scale representing an ESU or population, there may be many habitat actions within that area. Multiple treatment effects make it very difficult to assess the effects of specific actions on an ESU (see Section 2). Even though it may be impossible to assess specific treatment effects at larger spatial scales, it does not preclude one from conducting effectiveness monitoring at this scale. Indeed, one can assess the combined or cumulative effects of tributary actions on the ESU or population. However, additional effectiveness monitoring may be needed at finer scales to assess the effects of individual actions on the ESU or population.

Table 3. Relationship between biological indicators, spatial scales, and our ability to assess effects of specific management actions. Examples of each scale are shown in parentheses. Table is from Action Agencies/NOAA Fisheries RME Plan (2003).

Biological Indicators	**Example of spatial scales**	**Ability to assess effects of specific tributary actions**
ESU (Upper Columbia Spring Chinook) ↓	**Basins** (Upper Columbia) ↓	**Low**
Population (Wenatchee Spring Chinook) ↓	**Basin** (Wenatchee) ↓	↓
Sub-Population (Chiwawa River Spring Chinook) ↓	**Watershed** (Chiwawa River) ↓	
Local Group	**Reach** (5 km of the Chiwawa River)	**High**

Given the potential problems of multiple treatment effects, there are two general strategies for conducting effectiveness monitoring at different spatial scales. One strategy is a "project-based" approach, which addresses the effects of individual tributary projects at smaller spatial scales (e.g., stream or stream reach). This approach is identified in the Action Agencies/NOAA Fisheries Plan as the "Bottom-Up" approach. It is designed to assess the effects of specific projects in isolation of other tributary actions. That is, results from this type of effectiveness monitoring would not be confounded by actions occurring elsewhere in the basin. This approach requires that the investigator maintain control of all actions that occur within the assessment area (stream, watershed, or basin).

The second strategy is an "intensive" approach that addresses the cumulative effects of tributary actions at larger spatial scales (e.g., watershed or basin). This approach is identified in the Action Agencies/NOAA Fisheries Plan as the "Top-Down" Approach. The WSRFB (2003) refers to it as "Intensive (Validation) Monitoring." This approach requires intensive and extensive sampling of several indicator variables within the watershed or basin. Although the effects of individual projects on fish populations may not be assessed unequivocally, their cumulative effects can be measured.

Both approaches (project-based and intensive) require valid statistical and sampling designs. That is, both approaches require controls (reference conditions), replication, and probabilistic sampling. This plan recommends the use of BACI designs (see Section 2) with stratified random sampling (see Section 3) for both approaches. Both approaches will likely be implemented within the Wenatchee Basin.

SECTION 5: CLASSIFICATION

Both status/trend and effectiveness monitoring require landscape classification. The purpose of classification is to describe the "setting" in which monitoring occurs. This is necessary because biological and physical/environmental indicators may respond differently to tributary actions depending on landscape characteristics. A hierarchical classification system that captures a range of landscape characteristics should adequately describe the setting in which monitoring occurs. The idea advanced by hierarchical theory is that ecosystem processes and functions operating at different scales form a nested, interdependent system where one level influences other levels. Thus, an understanding of one level in a system is greatly informed by those levels above and below it.

A defensible classification system should include both ultimate and proximate control factors (Naiman et al. 1992). Ultimate controls include factors such as climate, geology, and vegetation that operate over large areas, are stable over long time periods, and act to shape the overall character and attainable conditions within a watershed or basin. Proximate controls are a function of ultimate factors and refer to local conditions of geology, landform, and biotic processes that operate over smaller areas and over shorter time periods. These factors include processes such as discharge, temperature, sediment input, and channel migration. Ultimate and proximate control characteristics help define flow (water and sediment) characteristics, which in turn help shape channel characteristics within broadly predictable ranges (Rosgen 1996).

This plan proposes a classification system that incorporates the entire spectrum of processes influencing stream features and recognizes the tiered/nested nature of landscape and aquatic features. This system captures physical/environmental differences spanning from the largest scale (regional setting) down to the channel segment (Table 4). The Action Agencies/NOAA Fisheries RME plan proposes the same classification system. By recording these descriptive characteristics, the investigator will be able to assess differential responses of indicator variables to proposed actions within different classes of streams and watersheds. Below I define each classification variable. Section 6 identifies recommended methods for measuring each variable.

Table 4. List of classification (stratification) variables that will be measured as part of monitoring within the Wenatchee Basin. The variables are nested according to spatial scale and their general characteristics. Table is from Action Agencies/NOAA Fisheries RME Plan (2003).

Spatial scale	General characteristics	Classification variable
Regional setting	Ecoregion	Bailey classification
		Omernik classification
	Physiography	Province
	Geology	Geologic districts
Drainage basin	Geomorphic features	Basin area
		Basin relief
		Drainage density
Valley segment	Valley characteristics	Valley bottom type
		Valley bottom width
		Valley bottom gradient
		Valley containment
Channel segment	Channel characteristics	Elevation
		Channel type (Rosgen)
		Bed-form type
		Channel gradient
	Riparian vegetation	Riparian cover group
		Riparian community type

As noted above, all watersheds that will be monitored will be classified according to their landscape characteristics. Table 4 lists the "core" set of classification variables. Section 6 provides a description of measurement protocols. Here I provide only a general description of each classification variable.

Regional Setting

Ecoregions:

Ecoregions are relatively uniform areas defined by generally coinciding boundaries of several key geographic variables. Ecoregions have been defined holistically using a set of physical and biotic factors (e.g., geology, climate, landform, soil, vegetation, and water). Of the systems available, this plan includes the two most commonly used ecoregion systems, Bailey (1978) and Omernik (1987). Bailey's approach uses macroclimate and prevailing plant formations to classify the continent into various levels of detail. Bailey's coarsest hierarchical classifications include domains, divisions, provinces, and sections. These regional classes are based on broad ecological climate zones and thermal and moisture limits for plant growth (Bailey 1998). Specifically, domains are groups of related climates, divisions are types of climate based on seasonality of precipitation or degree of dryness or cold, and provinces are based on macro features of vegetation. Provinces include characterizations of land-surface form, climate, vegetation, soils, and fauna. Sections are based on geomorphology, stratigraphy and lithology, soil taxa, potential natural vegetation, elevation, precipitation, temperature, growing season, surface water characteristics, and disturbance. Information from domains, divisions, and provinces can be used for modeling, sampling, strategic planning, and assessment. Information from sections can be used for strategic, multi-forest, statewide, and multi-agency analysis and assessment.

The system developed by Omernik (1987) is used to distinguish regional patterns of water quality in ecosystems as a result of land use. Omernik's system is suited for classifying aquatic ecoregions and monitoring water quality because of its ecological foundation, its level of resolution, and its use of physical, chemical, and biological information. Like Bailey's system, this system is hierarchical, dividing an area into finer regions in a series of levels. These levels are based on characterizations of land-surface form, potential natural vegetation, land use, and soils. Omernik's system has been extensively tested and found to correspond well to spatial patterns of water chemistry and fish distribution (Whittier et al. 1988).

Physiographic Province:

Physiographic province is the simplest division of a land area into hierarchical natural regions. In general, delineation of physiographic provinces is based on topography (mountains, plains, plateaus, and uplands) and, to a lesser extent, climate, which governs the processes that shape the landscape (weathering, erosion, and sedimentation). Specifically, provinces include descriptions of climate, vegetation, surficial deposits and soils, water supply or resources, mineral resources, and additional information on features particular to a given area (Hunt 1967). Physiographic provinces and drainage basins have traditionally been used in aquatic research to identify fish distributions (Hughes et al. 1987; Whittier et al. 1988).

Geology:

Geologic districts are areas of similar rock types or parent materials that are associated with distinctive structural features, plant assemblages, and similar hydrographic character. Geologic districts serve as ultimate controls that shape the overall character and attainable conditions within a watershed or basin. They are corollary to subsections identified in the U.S. Forest Service Land Systems Inventory (Wertz and Arnold 1972). Watershed and stream morphology are strongly influenced by geologic structure and composition (Frissell et al. 1986; Nawa et al. 1988). Structural features are the templates on which streams etch drainage patterns. The hydrologic character of landscapes is also influenced by the degree to which parent material has been weathered, the water-handling characteristics of the parent rock, and its weathering products. Like ecoregions, geologic districts do not change to other types in response to land uses.

Drainage Basin

Geomorphic Features:

This plan includes three important geomorphic features of drainage basins: basin area, basin relief, and drainage density. Basin area (a.k.a. drainage area or catchment area) is the total land area, measured in a horizontal plane, enclosed by a drainage divide, from which direct surface runoff from precipitation normally drains by gravity into a wetland, lake, or river. Basin relief is the difference in elevation between the highest and lowest points in the basin. It controls the stream gradient and therefore affects flood patterns and the amount of sediment that can be transported. Hadley and Schumm (1961) demonstrated that sediment load increases exponentially with basin relief. The last geomorphic feature, drainage density, is an index of the length of stream per unit area of basin and is calculated as the drainage area divided by the total stream length. This ratio represents the amount of stream necessary to drain the basin. High drainage density may indicate high water yield and sediment transport, high flood peaks, steep hills, and low suitability for certain land uses (e.g., agriculture).

Valley Segment

Valley Characteristics:

The plan incorporates four important features of the valley segment: valley bottom type, valley bottom width, valley bottom gradient, and valley confinement. Valley bottom types are distinguished by average channel gradient, valley form, and the geomorphic processes that shaped the valley (Cupp 1989a,b; Naiman et al. 1992). They correspond with distinctive hydrologic characteristics, especially the relationship between stream and alluvial ground water.[23] Valley bottom width is the ratio of the valley bottom[24] width to active channel width. Valley gradient is the slope or the change in vertical elevation per unit of horizontal valley distance. Valley gradient is typically measured in lengths of about 300 m (1000 ft) or more. Valley confinement refers to the degree that the valley walls confine the lateral migration of the stream channel. The degree of confinement can be classified as strongly confined (valley floor width < 2 channel widths), moderately confined (valley floor width = 2-4 channel widths), or unconfined (valley floor width > 4 channel widths).

Channel Segment

---

[23] Table 7.3 in Naiman et al. (1992) identifies and describes various valley bottom types.

[24] Valley bottom is defined as the essentially flat area adjacent to the stream channel.

Channel Characteristics:

The plan includes four important characteristics of the channel segment: elevation, channel gradient, channel type, and bed-form type. Elevation is the height of the stream channel above or below sea level. Channel gradient is the slope or the change in the vertical elevation of the channel per unit of horizontal distance. Channel gradient can be presented graphically as a stream profile.

Channel type follows the classification technique of Rosgen (1996) and is based on quantitative channel morphology indices.[25] These indices result in objective and consistent identification of stream types. The Rosgen technique consists of four different levels of classification. Level I describes the geomorphic characteristics that result from the integration of basin relief, landform, and valley morphology. Level II provides a more detailed morphological description of stream types. Level III describes the existing condition or "state" of the stream as it relates to its stability, response potential, and function. Level IV is the level at which measurements are taken to verify process relationships inferred from preceding analyses. All monitoring in the Wenatchee Basin will include at least Level I (geomorphic characterization) classification.

Bed-form type follows the classification proposed by Montgomery and Buffington (1993). This technique is comprehensive and is based on hierarchies of topographic and fluvial characteristics. This system provides a geomorphic, process-oriented method of identifying valley segments and stream reaches. It employs descriptors that are measurable and ecologically relevant. Montgomery and Buffington (1993) identified three valley segment types: colluvial, alluvial, and bedrock. They subdivided the valley types into one or more stream-reach types (bed-form types) depending on whether substrates are limited by the supply of sediment or by the fluvial transport of sediment. For example, depending on sediment supply and transport, Montgomery and Buffington (1993) recognized six alluvial bed-form types: braided, regime, pool/riffle, plane-bed, step-pool or cascade. Both colluvial and bedrock valley types consist of only one bed-form type. Only colluvial bed-forms occur in colluvial valleys and only bedrock bed-forms occur in bedrock valleys.

Riparian Vegetation:

Because riparian vegetation has an important influence on stream morphology and aquatic biota, the plan incorporates two characteristics of riparian vegetation: riparian cover group and riparian community type. Riparian cover group refers to the dominant vegetative cover type (Overton et al. 1997). The classification consists of two cover groups, wooded and meadow. Wooded riparian areas are characterized by streamside or upslope tree stands that have the potential to supply LWD to the stream channel. Meadow riparian areas are characterized by streamside or floodplain grasses, forbs, or shrubs (including willows) that have little potential to contribute LWD to the stream channel. Riparian community type is a

---

[25] Indices include entrenchment, gradient, width/depth ratio, sinuosity, and dominant channel material.

repeated and defined assemblage of riparian plant species. It requires knowledge of plant classification.

SECTION 6: SELECTION OF INDICATORS

In this section I identify the "core" set of biological and physical/environmental indicator variables that will be measured within all watersheds and streams that receive status/trend and effectiveness monitoring. The "core" list of variables represents the minimum, required variables that will be measured. Investigators may elect to measure additional variables depending on their objectives and past activities.

Indicator variables identified in this plan are consistent with those identified in the Action Agencies/NOAA Fisheries RME Plan. The Action Agencies/NOAA Fisheries selected indicators based on their review of the literature (e.g., Bjornn and Reiser 1991; Spence et al. 1996; Gregory and Bisson 1997; and Bauer and Ralph 1999) and several regional monitoring programs (e.g., PIBO, AREMP, EMAP, WSRFB, and the Oregon Plan). They selected variables that met various purposes including assessment of fish production and survival, identifying limiting factors, assessing effects of various land uses, and evaluating habitat actions. Their criteria for selecting variables were based on the following characteristics:

- Indicators should be sensitive to land-use activities or stresses.
- They should be consistent with other regional monitoring programs.
- They should lend themselves to reliable measurement.
- Physical/environmental indicators would relate quantitatively with fish production.

The indicators that the Action Agencies/NOAA Fisheries selected were consistent with most of the variables identified by the NMFS (1996) and USFWS (1998) as important attributes of "properly functioning condition." Indeed, the NMFS and USFWS use these indicators to evaluate the effects of land-management activities for conferencing, consultations, and permits under the ESA.

The indicators selected by the Action Agencies/NOAA Fisheries were also consistent with "key" parameters used in the Ecosystem Diagnosis and Treatment model. Recent analyses by Mobrand Biometrics indicated that certain physical/environmental parameters have a relatively important influence on modeled salmon production. These parameters included channel configuration, gradient, pool/riffle frequency, migration barriers, flow characteristics, water temperature, riparian function, fine sediment, backwater areas, and large woody debris (LWD) (K. Malone, Mobrand Biometrics, personal communication).

Below I identify and describe the "core" set of biological and physical/environmental variables that will be monitored in the Wenatchee Basin.

6.1 Biological Variables

The biological variables that will be measured in the Wenatchee Basin can be grouped into five general categories: adults, redds, parr, smolts, and macroinvertebrates. Each of these general categories consists of one or more indicator variables (Table 5). These biological indicators in concert will describe the characteristics of the populations of ESA-listed fish in the Wenatchee Basin and will provide information necessary for assessing recovery of listed stocks.

Table 5. Biological indicator variables to be monitored within the Wenatchee Basin.

General characteristics	Specific indicators
Adults	Escapement/Number
	Age structure
	Size
	Sex ratio
	Origin (hatchery or wild)
	Genetics
	Fecundity
Redds	Number
	Distribution
Parr/Juveniles	Abundance
	Distribution
	Size
Smolts	Number
	Size
	Genetics
Macroinvertebrates	Export of invertebrates

Adults

Escapement:

The plan includes escapement of mature adults as an important biological indicator of population health. Escapement is the total number of mature adults that enter or occur within a stream or watershed. Numbers of mature adults within a stream or watershed is a function of all the factors that affect the life history of the population.

Spawners:

The plan includes six indicators associated with the characteristics of the spawning populations: age structure, size, sex ratio, origin, genetics, and fecundity. Age structure describes the ages of adult fish within the spawning population. For anadromous species, age structure includes the number of years the fish spent in freshwater and number of years in salt water. Size describes the lengths and weights of adult fish within the spawning population. Sex ratio is the ratio of males to females within the spawning population. Origin identifies the parentage (hatchery or wild) of individuals within the spawning populations, while genetics defines not only the parentage but also within and between population variability. Fecundity is the number of eggs produced by a female.[26]

Redds

Abundance/Distribution:

Abundance describes the number of redds (nests) of ESA-listed fish species within the Wenatchee Basin. Total numbers will be estimated for ESA-listed anadromous species, while index counts will be made for bull trout. Distribution indicates the spatial arrangement and geographic extent of redds within the basin.

Parr

Abundance/Distribution:

Abundance describes the number of juvenile fish within specified stream reaches. Distribution is the spatial arrangement of juvenile fish within populations. It also captures the geographic range of individuals within the watershed or basin.

Condition:

The condition (or well-being) of fish can be assessed by measuring the length (fork length; FL) and weight of juvenile fish. The plan includes Fulton-type condition as the metric for well-being of juvenile fish (Anderson and Neumann 1996). The Fulton-type condition factor is of the form:

$$K_{FL} = (W/L^3) \times 100{,}000,$$

where K_{FL} = Fulton-type condition, W = weight in grams, and L = fork length in millimeters. The constant 100,000 is a scaling constant used to convert small decimals to mixed numbers so that the numbers can be more easily comprehended.

---

[26] By definition, *fecundity* refers to the number of eggs readied for spawning by a female. *Relative fecundity* is the number of eggs per unit of weight, while *total fecundity* is the number of eggs laid during the lifetime of the female. This plan refers to fecundity as the number of eggs per size (length and weight) of female.

Smolts

Abundance:

Abundance of smolts is an estimate of the total number of smolts produced within a watershed or basin. The estimate should be for an entire population or subpopulation.

Condition:

The Fulton-type condition factor describes the well-being of smolts within a population or subpopulation.

Genetics:

Genetic characterization (via DNA microsatellites) describes within- and between-population genetic variability of smolts.

Macroinvertebrates

Invertebrate Transport:

The plan includes export of invertebrates (aquatic and terrestrial) from headwaters to habitats downstream as an important attribute of productivity. The movement of prey items among habitats has a strong influence on fish populations, food webs, community dynamics and ecosystem processes (Wipfli and Gregovich 2002).

6.2 Physical/Environmental Variables

The physical/environmental variables that will be measured in the Wenatchee Basin can be grouped into seven general categories: water quality, habitat access, habitat quality, channel condition, riparian condition, flow/hydrology, and watershed condition. Each of these categories consists of one or more indicator variables (Table 6). In sum, these categories and their associated indicators address watershed process and "input" variables (e.g., artificial physical barriers, road density, and disturbance) as well as "outcome" variables (e.g., temperature, sediment, woody debris, pools, riparian habitat, etc.), as outlined in Hillman and Giorgi (2002) and the Action Agencies/NOAA Fisheries RME Plan.

What follows is a brief description of each physical/environmental indicator variable. Section 6 identifies recommended methods for measuring each indicator. Unless indicated otherwise, most of the information presented below has been summarized in Meehan (1991), MacDonald et al. (1991), Armantrout (1998), Bain and Stevenson (1999), OPSW (1999), Hillman and Giorgi (2002), and the Action Agencies/NOAA Fisheries RME Plan (2003).

Table 6. Physical/environmental indicator variables to be monitored within the Wenatchee Basin. Table is from Action Agencies/NOAA Fisheries RME Plan (2003).

General characteristics	Specific indicators
Water Quality	MWMT and MDMT
	Turbidity
	Depth fines
	pH
	DO
	Nitrogen
	Phosphorus
Habitat Access	Road crossings
	Diversion dams
	Fishways
Habitat Quality	Dominant substrate
	Embeddedness
	LWD (pieces/mile)
	Pools per mile
	Pool quality
	Side channels and backwaters
Channel condition	Width/depth ratio
	Wetted width
	Bankfull width
	Bank stability
Riparian Condition	Percent vegetation altered
Flows and Hydrology	Streamflow
Watershed Condition	Watershed road density
	Riparian-road index
	Equivalent clearcut area

Water Quality

Water Temperature:

The plan includes two temperature metrics that will serve as specific indicators of water temperature: maximum daily maximum temperature (MDMT) and maximum weekly maximum temperature (MWMT). MDMT is the single warmest daily maximum water temperature recorded during a given year or survey period. MWMT is the mean of daily maximum water temperatures measured over the warmest consecutive seven-day period.

MDMT is measured to establish compliance with the short-term exposure to extreme temperature criteria, while MWMT is measured to establish compliance with mean temperature criteria.

Sediment and Turbidity:

The plan includes two sediment-related specific indicators: turbidity and depth fines. Turbidity refers to the amount of light that is scattered or absorbed by a fluid. Suspended particles of fine sediments often increase turbidity of streams. However, other materials such as finely divided organic matter, colored organic compounds, plankton, and microorganisms can also increase turbidity of streams. Depth fines refer to the amount of fine sediment (<0.85 mm) within the streambed. Depth fines will be estimated at a depth between 6-12 inches within spawning gravels.

Contaminants and Nutrients:

The plan includes four specific indicators associated with contaminants and nutrients: pH, dissolved oxygen (DO), nitrogen, and phosphorus. Most of these indicators are commonly measured because of their sensitivity to land-use activities, municipal and industrial pollution, and their importance in aquatic ecosystems.

The plan included pH and DO because these parameters are often incorporated into water quality monitoring programs (e.g., OPSW 1999; Bilhimer et al. 2003). pH is defined as the concentration of hydrogen ions in water (moles per liter). It is a measure of how acidic or basic water is—it is not a measure of acidity or alkalinity (acidity and alkalinity are measures of the capacity of water to neutralize added base or acid, respectively). The logarithmic pH scale ranges from 0 to 14. Pure water has a pH of 7, which is the neutral point. Water is acidic if the pH value is less than 7 and basic if the value is greater than 7.

DO concentration refers to the amount of oxygen dissolved in water. Its concentration is usually measured in parts per million (ppm) or mg per liter (mg/L). The capacity of water to hold oxygen in solution is inversely proportional to the water temperature. Increased water temperature lowers the concentration of DO at saturation. Respiration (both plants and animals) and biochemical oxygen demand (BOD) are the primary factors that reduce DO in water. Photosynthesis and dissolution of atmospheric oxygen in water are the major oxygen sources.

The plan includes nitrogen and phosphorus as indicators of nutrient loading in streams. Nitrogen in aquatic ecosystems can be partitioned into dissolved and particulate nitrogen. Most water quality monitoring programs focus on dissolved nitrogen, because it is more readily available for both biological uptake and chemical transformations. Both dissolved and particulate nitrogen can be separated into inorganic and organic components. The primary inorganic forms are ammonia (NH_4^+), nitrate (NO_3^-), and nitrite (NO_2^-). Nitrate is the predominant form in unpolluted waters.

Phosphorus can also be separated into two fractions, dissolved and particulate. Dissolved phosphorus is found almost exclusively in the form of phosphate ions (PO_4^{-3}), which bind readily with other chemicals. There are three main classes of phosphate compounds: orthophosphates, condensed phosphates, and organically-bound phosphates. Each can occur as dissolved phosphorus or can be bound to particulate matter. In general, biota use only orthophosphates.

Habitat Access

Artificial Physical Barriers:

The plan includes three specific indicators associated with artificial physical barriers: road crossings (culverts), dams, and fishways. Roads and highways are common in the Wenatchee Basin and where they intersect streams they may block fish passage. Culverts can block passage of fish particularly in an upstream direction (WDFW 2000). In several cases, surveys have shown a difference in fish populations upstream and downstream from existing culverts, leading to the conclusion that free passage is not possible (Clay 1995). Dams and diversions that lack fish passage facilities can also block fish passage. Unscreened diversions may divert migrating fish into ditches and canals. Entrained fish can end in irrigated fields and orchards. Fishways are man-made structures that facilitate passage of fish through or over a barrier. Although these structures are intended to facilitate passage, they may actually impede fish passage (Clay 1995; WDFW 2000).

Habitat Quality

Substrate:

The Plan includes two specific indicators of substrate: dominant substrate and embeddedness. Dominant substrate refers to the most common particle size that makes up the composition of material along the streambed. This indicator describes the dominant material in spawning and rearing areas. Embeddedness is a measure of the degree to which fine sediments surround or bury larger particles. This measure is an indicator of the quality of over-wintering habitat for juvenile salmonids.

Large Woody Debris:

The plan includes the number of pieces of large woody debris (LWD) per stream kilometer as the one specific indicator of LWD in streams. LWD consists of large pieces of relatively stable woody material located within the bankfull channel and appearing to influence bankfull flows. LWD is also referred to as large organic debris (LOD) and coarse woody debris (CWD). The plan follows the definition of Armantrout (1998), who defined LWD as any piece of wood with a diameter greater than 10 cm and a length greater than 1 m. LWD can occur as a single piece (log), an aggregate (two or more clumped pieces, each of which qualifies as a single piece), or as a rootwad.

Pool Habitat:

The plan includes two specific indicators associated with pool habitat: number of pools per kilometer and pool quality. To be counted, a pool must span more than half the wetted width, be longer than it is wide, and include the thalweg. Pool quality refers to the ability of a pool to support the growth and survival of fish. Pool size (diameter and depth) and the amount and quality of cover determine overall pool quality. Pool cover is any material or condition that conceals or protects fish from predators or competitors and may consist of logs, organic debris, overhanging vegetation, cobble, boulders, undercut banks, or water depth.

Off-Channel Habitat:

Off-channel habitat consists of side-channels, backwater areas, alcoves or sidepools, off-channel pools, off-channel ponds, and oxbows. A side channel is a secondary channel that contains a portion of the streamflow from the main or primary channel. Backwater areas are secondary channels in which the inlet becomes blocked but the outlet remains connected to the main channel. Alcoves are deep areas along the shoreline of wide and shallow stream segments. Off-channel pools occur in riparian areas adjacent to the stream channels and remain connected to the channel. Off-channel ponds are not part of the active channel but are supplied with water from over bank flooding or through a connection with the main channel. These ponds are usually located on flood terraces and are called wall-based channel ponds when they occur near the base of valley walls. Finally, oxbows are bends or meanders in a stream that become detached from the stream channel either from natural fluvial processes or anthropogenic disturbances.

Channel Condition

Width/Depth Ratio:

The width/depth ratio is an index of the cross-section shape of a stream channel at bankfull level. The ratio is a sensitive measure of the response of a channel to changes in bank conditions. Increases in width/depth ratios, for example, indicate increased bank erosion, channel widening, and infilling of pools. Because streams almost always are several times wider than they are deep, a small change in depth can greatly affect the width/depth ratio.

Wetted Width:

Wetted width is the width of the water surface measured perpendicular to the direction of flow.

Bankfull Width:

Bankfull width is the width of the channel between the tops of the most pronounced banks on either side of a stream site or reach.

Streambank Condition:

The plan includes streambank stability as the one specific indicator of streambank condition. Streambank stability is an index of firmness or resistance to disintegration of a bank based on the percentage of the bank showing active erosion (alteration) and the presence of protective vegetation, woody material, or rock. A stable bank shows no evidence of breakdown, slumping, tension cracking or fracture, or erosion (Overton et al. 1997). Undercut banks are considered stable unless tension fractures show on the ground surface at the bank of the undercut.

Riparian Condition

Riparian Habitat:

The plan includes percent altered vegetation as the one specific indicator of riparian condition. Percent altered vegetation refers to the percentage of riparian vegetation along the stream channel that has been removed or altered by disturbance (includes both land-use activities and natural disturbances such as fires, floods, etc.).

Flows and Hydrology

Streamflows:

The plan includes three specific indicators of streamflows: change in peak flow, change in base flow, and change in timing of flow. Peak flow is the highest or maximum streamflow recorded within a specified period of time. Base flow is the streamflow sustained in a stream channel and is not a result of direct runoff. Base flow is derived from natural storage (i.e., outflow from groundwater, large lakes, or swamps), or sources other than rainfall. Timing of flow refers to the time when peak and base flows occur and the rate of rises and falls in the hydrograph. These indicators are based on "annual" flow patterns.

Watershed Conditions

Road Density:

A road is any open way for the passage of vehicles or trains. The plan includes both road density and the riparian-road index (RRI) as indicators of roads within watersheds. Road density is an index of the total miles of roads within a watershed. It is calculated as the total length of all roads (miles) within a watershed divided by the area of the watershed (miles^2). The RRI is expressed as the total mileage of roads within riparian areas divided by the total number of stream miles within the watershed (WFC 1998). For this index, riparian areas are defined as those falling within the federal buffers zones; that is, all areas within 300 ft of either side of a fish-bearing stream, within 150 ft of a permanent nonfish-bearing stream, or within the 100-year floodplain.

Watershed Disturbance:

The plan includes "equivalent clearcut area" (ECA) as the single indicator of watershed disturbance. ECA is defined as the area of a watershed that has been disturbed by timber harvest, roads, and fires, with an adjustment factor to account for the hydrologic recovery resulting from forest regeneration (USFS 1974; King 1989). The adjustment is based on regeneration (size of trees) and elevation.

6.3 Recommended Indicators

As noted earlier, the biological and physical/environmental indicators identified in this section represent a "core" list of variables that will be measured in the Wenatchee Basin. This plan does not preclude the investigator from measuring other indicator variables. Which variables will be measured depends on the type of monitoring (status/trend vs. effectiveness), the target fish species, and the type of tributary action implemented. Below I identify the appropriate indicators for each type of monitoring.

Effectiveness Monitoring—This plan does not recommend that all the indicators listed in Tables 5 and 6 be measured for each tributary action. Different biological indicators will be measured depending on the fish species of interest (Table 7). All biological indicators identified in Table 5 will be measured for actions that affect anadromous species (spring chinook, summer/fall chinook, steelhead, and sockeye salmon). For resident species (bull trout and cutthroat trout), however, indicators related to smolts and origin will not be measured.

The plan recommends that only those physical/environmental indicators that are linked directly to the proposed action be measured. In other words, the most useful indicators are likely to be those that represent the first links of the cause-and-effect chain. Because different projects have different objectives and desired effects, the investigator only needs to measure those indicators directly influenced on the chain of causality between the habitat action and the effect (Table 8). This approach differs from the Action Agencies/NOAA Fisheries Plan, which requires all indicators be measured, regardless of the type of habitat action implemented.

Status/Trend Monitoring—All the physical/environmental indicators identified in Table 6 will be measured as part of status/trend monitoring in the Wenatchee Basin. In contrast, different biological indicators will be measured depending on the target fish species (Table 7). As with effectiveness monitoring, all biological indicators identified in Table 5 will be measured for anadromous species. Indicators related to smolts and origin will not be measured for resident species.

Table 7. Biological indicator variables that will be measured (marked with an "X") for anadromous (spring chinook, summer/fall chinook, steelhead, and sockeye salmon) and resident (bull trout and cutthroat trout) fish species during status/trend and effectiveness monitoring in the Wenatchee Basin.

General characteristics	Specific indicators	Anadromous species	Resident species
Adults	Escapement/Number	X	X
	Age structure	X	X
	Size	X	X
	Sex ratio	X	X
	Origin (hatchery or wild)	X	
	Genetics	X	X
	Fecundity	X	
Redds	Number	X	X
	Distribution	X	X
Parr/Juveniles	Abundance	X	X
	Distribution	X	X
	Size	X	X
Smolts	Number	X	
	Size	X	
	Genetics	X	
Macroinvertebrates	Export of invertebrates	X	X

Table 8. Rankings of the usefulness of physical/environmental indicators to monitoring effects of different tributary habitat actions. Rankings vary from 1 = highly likely to be useful; 2 = moderately likely to be useful; and 3 = unlikely to be useful or little relationship, although the indicator may be useful under certain conditions or may help interpret data from a primary indicator. Table is modified from Hillman and Giorgi (2002). The different classes of habitat actions are from the Action Agencies/NOAA Fisheries RME Plan.

General characteristics	**Specific indicators**	**Different classes of habitat actions**							
		Diversion screens	**Barrier removal**	**Sediment reduction**	**Water quality improvement**	**Nutrient enhancement**	**Instream flows**	**Riparian habitat**	**Instream structure**
Water quality	MWMT/MDMT	3	2	3	1	2	1-2	1	3
	Turbidity	3	1-2	1	1	1	1-2	2	3
	Depth fines	3	1-2	1	1-2	2	2	2	1-2
	pH	3	3	3	1	1	3	2-3	3
	DO	3	2-3	2-3	1	1	1-2	2-3	3
	Nitrogen	3	3	3	1	1	3	2	3
	Phosphorus	3	3	3	1	1	3	2	3
Habitat access	Road crossings	3	1	3	3	3	3	3	3
	Diversion dams	1-2	1	3	3	3	2	3	3
	Fishways	2-3	1	3	3	3	3	3	3
Habitat quality	Dominant substrate	3	2	1	3	3	1-2	2	1-2
	Embeddedness	3	1-2	1	1-2	3	1-2	2	1-2
	LWD	3	3	3	3	3	2	1	1
	Pool frequency	3	1-2	1-2	3	3	1-2	1-2	1
	Pool quality	3	1-2	1	2	3	1	1-2	1
	Off-channel habitat	3	2	2	3	3	1	1-2	1
Channel condition	Width/depth	3	1-2	1-2	3	3	1-2	1-2	1
	Wetted width	3	1-2	1-2	3	3	1-2	1-2	1
	Bankful width	3	1-2	1-2	3	3	1-2	1-2	1
	Bank stability	3	2	1-2	3	3	2	1	1
Riparian condition	Percent veg altered	3	3	2	3	3	2	1	1-2
Flows/hydrology	Streamflows	3	1-2	3	3	3	1	2	1-2
Watershed condition	Road density	3	3	1-2	2	3	2-3	2-3	2
	Riparian-road index	3	3	1-2	2	3	2-3	1	2
	Equivalent clearcut	3	3	1-2	2	3	1	2	2

SECTION 7: MEASURING PROTOCOLS

An important component of the regional monitoring strategies (ISAB, Action Agencies/NOAA Fisheries, and WSRFB) is that they all recommend that the same measurement method be used to measure a given indicator. The reason for this is to allow comparisons of biological and physical/environmental conditions within and among watersheds and basins. In this section I identify methods to be used to measure biological and physical/environmental indicators. The methods identified in this plan are consistent with those described in the Action Agencies/NOAA Fisheries RME Plan.

The Action Agencies/NOAA Fisheries monitoring group reviewed several publications, including the work of Johnson et al. (2001) that describe methods for measuring indicators. Not surprisingly, there can be several different methods for measuring the same variable. For example, channel substrate can be described using surface visual analysis, peddle counts, or substrate core samples (either McNeil core samples or freeze-core samples). These techniques range from the easiest and fastest to the most involved and informative. As a result, one can define two levels of sampling methods. Level 1 (extensive methods) involves fast and easy methods that can be completed at multiple sites, while Level 2 (intensive methods) includes methods that increase accuracy and precision but require more sampling time. The Action Agencies/NOAA Fisheries monitoring group selected primarily Level 2 methods, which minimize sampling error.

Before I identify measuring protocols, it is important to define a few terms. These terms are consistent with the Action Agencies/NOAA Fisheries RME Plan.

Reach (effectiveness monitoring) – for effectiveness monitoring, a stream reach is defined as a relatively homogeneous stretch of a stream having similar regional, drainage basin, valley segment, and channel segment characteristics and a repetitious sequence of habitat types. Reaches are identified by using a list of classification (stratification) variables (from Table 4). Reaches may contain one or more sites. The starting point and ending point of reaches will be measured with Global Positioning System (GPS) and recorded as Universal Transverse Mercator (UTM).

Reach (status/trend monitoring) – for status/trend monitoring, a reach is a length of stream (20 times the average bankfull width, but not less than 150 m long)[27] selected with a systematic randomized process (GRTS design). GRTS selects a point on the "blue-line" stream network represented on a USGS map. This point is referred to as the "X-site." The X-site identifies the midpoint of the reach. That is, the sampling reach extends a distance of 10 times the average bankfull width upstream and downstream from the X-site. Biological and physical/environmental indicators are measured within

---

[27] This reach length differs from the EMAP protocol, which recommends a reach length of 40 times the average channel width (Lazorchak et al. 1998). The use of 20 times the average bankfull width is consistent with the Action Agencies/NOAA Fisheries RME Plan and the length of effectiveness monitoring sites.

the reach. The X-site and the upstream and downstream ends of the reach will be measured with GPS and recorded as UTM.

Site (effectiveness monitoring) – a site is an area of the effectiveness monitoring stream reach that forms the smallest sampling unit with a defined boundary. Site length depends on the width of the stream channel. Sites will be 20 times the average bankfull width with a minimum length of 150 m and a maximum length of 500 m. The upstream and downstream boundaries of the site will be measured with GPS and recorded as UTM.

Transect – a transect is a straight line across a stream channel, perpendicular to the flow, along which habitat features such as depth or substrate are measured at pre-determined intervals. Effectiveness monitoring sites and status/trend monitoring reaches will be divided into 11 evenly-spaced transects by dividing the site into 10 equidistant intervals with "transect 1" at the downstream end of the site or reach and "transect 11" at the upstream end of the site or reach.

7.1 Classification Variables

As indicated in Section 5, all watersheds that will be monitored will be classified according to their landscape characteristics. Table 9 identifies classification variables and recommended protocols for measuring them. Because time and space do not allow me to describe methods in detail, I only identify recommended methods and instruments. I refer the reader to the cited documents for detailed descriptions of methods and measuring instruments.

Regional Setting

Ecoregions:

The plan includes the two most commonly used ecoregion systems, Bailey (1978) and Omernik (1987). Until there is a better understand of the relationships between fish abundance and distribution and the two classes of ecoregions, investigators should use both classifications (Bailey's and Omernik's). Chapter 3 in Bain and Stevenson (1999) outlines protocols for describing ecoregions. Published maps of ecoregions are available to assist with classification work. This work will be updated once every 20 years.

Physiographic Province:

Investigators will describe physiographic provinces for all watersheds that will be monitored. Chapter 3 in Bain and Stevenson (1999) outlines methods for describing this variable. Physiographic maps are available to aid classification work. Investigators will update physiographic provinces once every 20 years.

Table 9. List of classification (stratification) variables, their corresponding measurement protocols, and temporal sampling frequency. Table is from the Action Agencies/NOAA Fisheries RME Plan.

Spatial scale	General characteristics	Classification variable	Recommended protocol	Sampling frequency (years)
Regional setting	Ecoregion	Bailey classification	Bain and Stevenson (1999)	20
		Omernik classification	Bain and Stevenson (1999)	20
	Physiography	Province	Bain and Stevenson (1999)	20
	Geology	Geologic districts	Overton et al. (1997)	20
Drainage basin	Geomorphic features	Basin area	Bain and Stevenson (1999)	20
		Basin relief	Bain and Stevenson (1999)	20
		Drainage density	Bain and Stevenson (1999)	20
Valley segment	Valley characteristics	Valley bottom type	Cupp (1989); Naiman et al. (1992)	20
		Valley bottom width	Naiman et al. (1992)	20
		Valley bottom gradient	Naiman et al. (1992)	20
		Valley containment	Bisson and Montgomery (1996)	20
Channel segment	Channel characteristics	Elevation	Overton et al. (1997)	10
		Channel type (Rosgen)	Rosgen (1996)	10
		Bed-form type	Bisson and Montgomery (1996)	10
		Channel gradient	Overton et al. (1997)	10
	Riparian vegetation	Riparian cover group	Overton et al. (1997)	5
		Riparian community type	Overton et al. (1997)	5

Geology:

Geologic districts are areas of similar rock types or parent materials that are associated with distinctive structural features, plant assemblages, and similar hydrographic character. Geologic districts can be identified following the methods described in Overton et al. (1997). Published geology maps aid in the classification of rock types. This work will be updated once every 20 years.

Drainage Basin

Geomorphic Features:

Basin area, basin relief, and drainage density describe the geomorphic features of a watershed. Chapter 4 in Bain and Stevenson (1999) outlines standard methods for estimating these parameters. Investigators will use USGS topographic maps (1:24,000 scale) and GIS to estimate these parameters. This work will be updated once every 20 years.

Valley Segment

Valley Characteristics:

The plan includes four important features of the valley segment: valley bottom type, valley bottom width, valley bottom gradient, and valley confinement. Investigators will follow the methods of Cupp (1989a,b) and Naiman et al. (1992) to describe valley bottom types. Naiman et al. (1992) describes methods for measuring valley bottom width and valley bottom gradient. Bisson and Montgomery (1996) outline methods for measuring valley confinement. GIS will aid in estimating these parameters. These variables will be updated once every 20 years.

Channel Segment

Channel Characteristics:

The plan includes four characteristics of the channel segment: elevation, channel gradient, channel type, and bed-form type. Each of these characteristics will be measured within each watershed that will be monitored. Overton et al. (1997) describe methods for measuring elevation and channel gradient. Bisson and Montgomery (1996) describe in detail the method for identifying channel bed-form types, while Rosgen (1996) describes methods for classifying channel types. All classification work will include at least Level I (geomorphic characterization) channel type classification. Depending on the objectives of the monitoring program, additional levels of classification may be necessary. These variables will be updated once every 10 years.

Riparian Vegetation:

The Plan includes two characteristics of riparian vegetation: riparian cover group and riparian community type. Investigators will use the methods described in Overton et al. (1997) to assess cover group and riparian community classification.

7.2 Biological Indicators

This section identifies the methods and instruments that should be used to measure biological indicators. Table 10 identifies indicator variables, recommended protocols, and sampling frequency. I refer the reader to the cited documents for a more detailed description of each method.

Table 10. Recommended protocols and sampling frequency for biological indicator variables.

General characteristics	Specific indicators	Recommended protocol	Sampling frequency
Adults	Escapement/Number	Dolloff et al. (1996); Reynolds (1996); Van Deventer and Platts (1989)	Annual
	Age structure	Borgerson (1992)	Annual
	Size	Anderson and Neumann (1996)	Annual
	Sex ratio	Strange (1996)	Annual
	Origin (hatchery or wild)	Borgerson (1992)	Annual
	Genetics	WDFW Genetics Lab	Annual
	Fecundity	Cailliet et al. (1986)	Annual
Redds	Number	Mosey and Murphy (2002)	Annual
	Distribution	Mosey and Murphy (2002)	Annual
Parr/Juveniles	Abundance/Distribution	Dolloff et al. (1996); Reynolds (1996); Van Deventer and Platts (1989)	Annual
	Size	Anderson and Neumann (1996)	Annual
Smolts	Number	Murdoch et al. (1999)	Annual
	Size	Anderson and Neumann (1996)	Annual
	Genetics	WDFW Genetics Lab	Annual
Macroinvertebrates	Export of invertebrates	Wipfli and Gregovich (2002)	Annual

Adults

Escapement:

The plan includes escapement/number of mature adults as an important biological indicator of population health. Escapement of anadromous fish into the Wenatchee Basin can be estimated roughly as the difference between fish counts at Rock Island Dam and Rocky Reach Dam (with some correction for fallback). Counts at dams should be made with video operated continuously during the upstream migration of anadromous salmonids. Counts of adults at weirs are more accurate and should be used whenever possible. This method is recommended if accurate estimates of escapements into specific watersheds are necessary.

Numbers of resident adult fish should be estimated within status/trend monitoring reaches and effectiveness monitoring sites using underwater observations (snorkeling) and electrofishing surveys. Snorkeling, which is a quick, nondestructive method that is not restricted by deep water and low conductivities[28], is the "primary" sampling method in this plan. Snorkel

---

[28] Hillman and Miller (2002) reported that snorkel estimates were more accurate than electrofishing estimates in the Chiwawa River, because low conductivity (35 µmhos) in the river reduced the efficiency of electrofishing. They noted that electrofishing estimates were at best 68% of snorkel estimates.

surveys will follow the protocols identified in Dolloff et al. (1996). Accurate estimates of adult bull trout may require nighttime snorkeling. However, Hillman and Chapman (1996) counted more adult bull trout during the day than at night in the Blackfoot River, Montana, because adult bull trout were unable to conceal themselves, making them readily visible to snorkelers. Both daytime and nighttime surveys should be conducted for at least two years to see which survey time (daytime or nighttime) provides the best estimate of resident adult fish.

Electrofishing is the "secondary" method and will be used within a sub-sample of snorkel sites. The plan recommends that at least two randomly-selected sites within each watershed[29] be sampled with both snorkeling and electrofishing.[30] The purpose for this is to establish a relationship between the methods and to collect fish for assessment of condition (length and weight), age, gender, and genetics. Electrofishing will follow the protocols outlined in Reynolds (1996). This plan recommends the removal-depletion method of electrofishing. Population numbers and 95% confidence intervals will be estimated with the maximum-likelihood formula (Van Deventer and Platts 1989).

Spawners:

The plan includes six indicators associated with the characteristics of the spawning populations: age structure, size, sex ratio, origin, genetics, and fecundity. For anadromous fish, most of these characteristics will be collected from live fish trapped at weirs or from carcasses sampled during spawning surveys. Scales will be pulled from live fish and carcasses. Scales will be read to determine age structure and origin (wild or hatchery). Presence or absence of an adipose fin will also determine origin. Age analysis will be completed by methods described by Borgerson (1992). Size will be reported as both fork length (anterior tip to the median caudal fin rays) and hypural length (mid-eye to hypural plate) (Anderson and Neumann 1996). The latter is necessary because some carcasses will have decomposed to a point where fork length cannot be measured accurately. The gender of each fish sampled will be recorded (Strange 1996). Fecundity (total number of eggs produced by a given size female) will be estimated for fish collected for hatchery broodstock and from dead pre-spawn females collected during spawning surveys (Cailliet et al. 1986). Finally, genetic samples will be collected and analyzed according to the protocols being refined at the WDFW Genetics Lab. All sampled carcasses will be marked to avoid resampling.

Many of the characteristics identified above for anadromous fish will be collected from resident fish during electrofishing surveys, collection at weirs, and during spawning surveys. Characteristics such as age structure (from scales), size (fork length and mid-eye to hypural length; mm), weight (g), and genetic samples can be collected from adults trapped at weirs and during electrofishing surveys. Gender can be recorded for those fish found dead during spawning surveys. The protocols identified above can be used to measure characteristics of resident fish.

---

[29] In the Wenatchee Basin there are nine major watersheds. If two sites are sampled within each watershed, the investigator would sample a total of 18 sites.

[30] Sampling within a site should occur within the same day and sites should be blocked to prevent movement into and out of the site during and between sampling.

Redds

Abundance/Distribution:

This plan includes abundance and distribution of salmonid redds as indicators of population health. The plan calls for a complete census of anadromous fish redds and a probabilistic sample of resident fish redds. Throughout the spawning period, investigators will conduct weekly redd surveys following the methods of Mosey and Murphy (2002). Each week new redds will be counted, mapped, and marked.[31] Marking is needed to avoid recounting redds during subsequent surveys. The entire distribution of anadromous spawning areas will be sampled. For resident fish, index areas selected according to a probabilistic sampling design (e.g., stratified random sampling) will be surveyed for redds.

Parr

Abundance/Distribution:

The plan includes the abundance and distribution of juvenile fish as an indicator of population health. Juvenile numbers will be estimated with snorkeling and electrofishing within status/trend monitoring reaches and effectiveness monitoring sites. Snorkeling is the "primary" sampling method in this plan and will follow the protocols identified in Dolloff et al. (1996). Accurate estimates of juvenile bull trout may likely require nighttime snorkeling. Therefore, both daytime and nighttime surveys should be conducted for at least two years to see which survey time (daytime or nighttime) provides the best estimate of juvenile fish.

Electrofishing is the "secondary" method and will be used within at least two randomly-selected sites within each watershed (same sites used to sample adult fish). Electrofishing will follow the protocols outlined in Reynolds (1996). This plan recommends the removal-depletion method of electrofishing. Population numbers and 95% confidence intervals will be estimated with the maximum-likelihood formula (Van Deventer and Platts 1989).

Juvenile fish collected during electrofishing will be measured (see below) and at least 5,000 juvenile chinook and 5,000 steelhead will be implanted with PIT tags. The sample size of 5,000 for anadromous populations in the Upper Columbia Basin was estimated by the Action Agencies/NOAA Fisheries Monitoring Group. This is the minimum number needed to estimate life-stage survival rates.

---

[31] Because of inclement weather and high streamflows, surveys for steelhead redds may not be made on regularly timed intervals. Adjusting surveys to fit environmental conditions may be necessary.

Condition:

The plan includes Fulton-type condition as the metric for well-being of juvenile fish. Juvenile fish collected during electrofishing and with rotary traps will be measured (fork length; mm) and weighed (g). Fulton-type condition will be estimated with methods described in Anderson and Neumann (1996).

Smolts

Abundance:

Abundance of smolts is an estimate of the total number of smolts produced within a watershed or basin. Investigators will use floating screw traps to collect downstream migrating smolts. Traps will operate for at least the entire period of the smolt migration. Trapping efficiency, based on mark/recapture will be estimated throughout the trapping period. Methods for operating the trap, estimating efficiency, and the frequency at which efficiency tests are conducted are described in Murdoch et al. (1999).

Condition:

The Fulton-type condition factor describes the well-being of smolts within a population or subpopulation. Smolts collected with rotary traps will be measured (fork length; mm) and weighed (g). Fulton-type condition will be estimated with methods described in Anderson and Neumann (1996).

Genetics:

Genetic characterization (via DNA microsatellites) describes within- and between-population genetic variability of smolts. DNA samples from a systematic sample of smolts[32] will be collected and analyzed according to the protocols being refined at the WDFW Genetics Lab.

Macroinvertebrates

Invertebrate Transport:

The plan includes export of invertebrates (aquatic and terrestrial) from headwaters to habitats downstream as an attribute of freshwater productivity. Investigators will follow the methods described in Wipfli and Gregovich (2002) to assess energy sources for downstream food webs. The method requires the placement of sampling stations near tributary junctions of fishless and fish-bearing streams. Specially-modified drift nets (Wipfli and Gregovich 2002) will capture invertebrates and particulate organic matter. This work will be conducted annually during base-flow conditions.

---

[32] The total number of smolts needed to characterize within and between-population genetic variability is presently unknown. Therefore, "**k**" (i.e., the k^{th} smolt sampled) remains undefined.

7.3 Physical/Environmental Indicators

In this section I identify the methods and instruments needed to measure physical/environmental indicators. Table 11 identifies indicator variables, recommended protocols for measuring indicators, and sampling frequency. There is no space here to describe each method in detail; therefore, I refer the reader to the cited documents for detailed descriptions of methods and measuring instruments. Importantly, all habitat sampling would follow fish sampling (snorkeling and electrofishing) within status/trend monitoring reaches and effectiveness monitoring sites.

Water Quality

Water Temperature:

The plan includes two temperature metrics that will serve as specific indicators of water temperature: maximum daily maximum temperature (MDMT) and maximum weekly maximum temperature (MWMT). Data loggers will be used to measure MWMT and MDMT. Zaroban (2000) describes pre-placement procedures (e.g., selecting loggers and calibration of loggers), placement procedures (e.g., launching loggers, site selection, logger placement, and locality documentation), and retrieval procedures. This manual provides standard methods for conducting temperature-monitoring studies associated with land-management activities and for characterizing temperature regimes throughout a watershed.

The number of loggers used will depend on the number of reaches and treatment and control sites. For effectiveness monitoring, at a minimum, at least one logger will measure water temperatures at the downstream end and one at the upstream end of each reach that contains treatment or control sites. Additional measurements may be needed within reaches (at treatment sites) if management actions directly affect water temperature (e.g., restore riparian function). For status/trend monitoring, one logger will be placed at or near the X-site within the monitoring reach. Temperatures will be monitored continuously throughout the year.

Table 11 Recommended protocols and sampling frequency of physical/environmental indicator variables. Table is modified from Action Agencies/NOAA Fisheries RME Plan.

General characteristics	**Specific indicators**	**Recommended protocols**	**Sampling frequency**
Water Quality	MWMT/MDMT	Zaroban (2000)	Continuous
	Turbidity	OPSW (1999)	High and base-flow periods (2 times/yr)
	Depth fines	Schuett-Hames (1999)	Annual
	pH	OPSW (1999)	High and base-flow periods (2 times/yr)
	DO	OPSW (1999)	High and base-flow periods (2 times/yr)
	Nitrogen	OPSW (1999)	High and base-flow periods (2 times/yr)
	Phosphorus	OPSW (1999)	High and base-flow periods (2 times/yr)
Habitat Access	Road crossings	Parker (2000); WDFW (2000)	Annual
	Diversion dams	WDFW (2000)	Annual
	Fishways	WDFW (2000)	Annual
Habitat Quality	Dominant substrate	Bunte & Abt (2001); Platts et al. (1983)	Annual
	Embeddedness	MacDonald et al. (1991)	Annual
	LWD (pieces/mile)	BURPTAC (1999)	Annual
	Pools per mile	Overton et al. (1997)	Annual
	Pool quality	Platts et al. (1983)	Annual
	Off-channels habitats	WFPB (1995)	Annual
Channel condition	Width/depth ratio	BURPTAC (1999)	Annual
	Wetted width	Bain & Stevenson (1999)	Annual
	Bankfull width	Bain & Stevenson (1999)	Annual
	Bank stability	BURPTAC (1999)	Annual
Riparian Condition	Percent veg altered	Platts et al. (1987)	Annual
Flows and Hydrology	Streamflow	Bain & Stevenson (1999)	Continuous
Watershed Condition	Watershed road density	WFC (1998); Reeves et al. (2001)	Annual
	Riparian-road index	WFC (1998)	Annual
	Equivalent clearcut area	USFS (1974); King (1989)	Annual

Sediment and Turbidity:

The plan includes two sediment-related specific indicators: turbidity and depth fines. Investigators will measure turbidity with a portable turbidimeter (calibrated on the nephelometric turbidity method) following protocols described in Chapter 11 in OPSW (1999). This guidebook provides a standardized method for measuring turbidity, data quality guidelines, equipment, field measurement procedures, and methods to store and analyze turbidity data. For effectiveness monitoring, at a minimum, turbidity will be measured at the downstream end and at the upstream end of each reach that contains treatment or control sites. Additional measurements may be needed at treatment sites within reaches if management actions directly affect turbidity (e.g., sediment reduction actions). For status/trend monitoring, turbidity should be measured at or near the X-site within the monitoring reach. Turbidity will be measured during high flow (spring) and base-flow (summer) conditions.

Investigators will measure depth fines with McNeil core samplers following methods described in Schuett-Hames et al. (1999). For effectiveness monitoring, three randomly-selected samples (subsamples) will be taken from each spawning area (pool tailout or riffle) within each site (samples will not be taken from sites that lack spawning areas). For status/trend monitoring, three subsamples from one randomly-selected spawning area within a reach will be collected. The volumetric method will be used for processing samples sorted via a standard set of sieves. The volumetric method measures the millimeters of water displaced by particles of different size classes. At a minimum, the following sieves will be used to sort particles: 64.0 mm, 16.0 mm, 6.4 mm, 4.0 mm, 1.0 mm, 0.85 mm, 0.50 mm, 0.25 mm, and 0.125 mm. Fines will be measured once annually during base-flow conditions.

Contaminants and Nutrients:

The plan includes four specific indicators associated with contaminants and nutrients: pH, dissolved oxygen (DO), nitrogen, and phosphorus. The procedures described by OPSW (1999) will be used to measure pH, dissolved oxygen, nitrogen, and phosphorus. The guidebook provides a standardized method for measuring pH (pH meter—Chapter 8), DO (Winkler Titration Method—Chapter 7)[33], and nitrate/nitrites, Kjeldahl nitrogen, total phosphorous, and orthophosphates (Chapter 10), including criteria for data quality guidelines, equipment, field measurement procedures, and methods to store and analyze water quality data. For effectiveness monitoring, at a minimum, these indicators will be measured at the downstream end and upstream end of each reach that contains treatment or controls sites. Additional measurements may be needed at treatment sites within reaches if management actions directly affect these water-quality parameters (e.g., nutrient enhancement). For status/trend monitoring, samples should be collected at or near the X-site within the monitoring reach. These indicators will be measured once during high flow (spring) and during base flow (summer).

---

[33] According to OPSW (1999), the Winkler Titration Method is the most accurate method for measuring DO concentration. Although this plan recommends the Winkler Titration Method, calibrated DO meters with an accuracy of ±0.2 mg/L can be used in place of the chemical method.

Habitat Access:

Artificial Physical Barriers:

The plan includes three specific indicators associated with artificial physical barriers: road crossings (culverts), dams, and fishways. The WDFW (2000) manual provides guidance and methods on how to identify, inventory, and evaluate culverts, dams, and fishways that impede fish passage. WDFW (2000) also provides methods for estimating the potential habitat gained upstream from barriers, allowing prioritization of restoration projects. The manual by Parker (2000) focuses on culverts. The methods outlined in this manual assess connectivity of fish habitats on a watershed scale. These manuals will be used to identify all fish passage barriers within monitoring reaches. Assessment of fish passage barriers will occur once annually during base-flow conditions.

Habitat Quality

Substrate:

The plan includes two specific indicators of substrate: dominant substrate and embeddedness. Pebble counts will be used to identify substrate composition. Investigators will measure substrate at five equidistant points along each of the 11 transects. Following Bunte and Abt (2001), a 60 x 60-cm sampling frame will be used to sample substrate at each point along a transect. The sampling frame will be divided into four grid points by spacing the elastic bands 30 cm from each other. The sampling frame is intended to reduce operator influence on the selection of particle sizes. In field tests, the sampling frame produced slightly coarser size distributions than the traditional heel-to-toe walk. The sampling frame also produced more similar sampling results between two investigators than heel-to-toe walks. Classification of bed material by particle size will follow Table 4 in Platts et al. (1983).

Investigators will follow methods described in MacDonald et al. (1991) for measuring embeddedness. The method involves the use of a 60-cm-diameter hoop as the basic sample unit. The use of hoops rather than individual particles as the basic sampling unit substantially increases the number of particles that must be measured, but reduces the variability among sample units. This makes it easier to detect change and results in an embeddedness value that more closely represents the condition of the stream reach. Embeddedness will be collected within riffles in the lower and upper portions of each sampling reach. For effectiveness monitoring, additional sampling will occur within sites that are treated with actions that directly affect embeddedness (e.g., sediment reduction activities). For status/trend monitoring, embeddedness will be measured in one randomly selected riffle within the monitoring reach. Substrate indicators will be measured once annually during base-flow stream conditions.

Large Woody Debris:

Large woody debris (LWD) consists of large pieces of relatively stable woody material located within the bankfull channel and appearing to influence bankfull flows. Investigators will follow methods described in BURPTAC (1999) for estimating the number of pieces of large woody debris in forested streams. The guidelines describe procedures for dealing with single pieces and aggregates. Pieces of LWD will be counted in all monitoring reaches within forested streams. This indicator will be measured once annually during base-flow conditions.

Pool Habitat:

The plan includes two indicators associated with pool habitat: number of pools per mile and pool quality. Investigators will count the number of pools throughout a monitoring reach. To be counted, a pool must span more than half the wetted width, be longer than it is wide, and include the thalweg. Overton et al. (1997) provide a good description of the various types of pools and how to identify them. Pool frequency will be measured in all monitoring reaches.

Platts et al. (1983) describe methods for estimating pool quality (see their Table 1). This plan includes a slight modification to the Platts protocol by adding residual pool depth to the criteria. Residual pool depth is the difference between the maximum pool depth and the pool crest outlet depth (Overton et al. (1997) describe methods for measuring these two depths). Residual pool depth is independent of streamflow at time of measurement and is sensitive to land-management actions. For effectiveness monitoring, pool quality will be assessed for all pools within treatment and control sites. For status/trend monitoring, pool quality will be measured for all pools within a reach. Both pool frequency and pool quality will be measured once annually during base-flow conditions.

Off-Channel Habitat:

Off-channel habitat consists of side-channels, backwater areas, alcoves or sidepools, off-channel pools, off-channel ponds, and oxbows. Following the definitions for each off-channel habitat type (see Section 6), the investigator will enumerate the number of each type of off-channel habitat within a monitoring reach. This indicator is specific only to channels with gradients <3% (WFPB 1995). Sampling will occur once annually during base-flow conditions.

Channel Condition

Width/Depth Ratio:

The width/depth ratio is an index of the cross-section shape of a stream channel at bankfull level. The ratio is expressed as bankfull width (geomorphic term) divided by the mean cross-section depth. To measure width/depth ratio, the investigator will follow the protocol described in BURPTAC (1999), with one exception. Rather than measure wetted width and wetted depth, the investigator will measure mean bankfull width and mean bankfull depth. BURPTAC (1999) describes methods for estimating W/D rations in both single channels and split channels. This indicator will be measured at the 11 evenly-spaced transects within each reach (for status/trend monitoring) or treatment and control site (for effectiveness monitoring). Sampling will occur once annually during base-flow conditions.

Wetted Width:

Wetted width is the width of the water surface measured perpendicular to the direction of flow. Wetted width will be measured at the 11 evenly-spaced transects within each reach (for status/trend monitoring) or treatment and control site (for effectiveness monitoring) following the protocol in Bain and Stevenson (1999). Widths of multiple channels are summed to represent the total wetted width. Sampling will occur once annually during base-flow conditions.

Bankfull Width:

Bankfull width is the width of the channel between the tops of the most pronounced banks on either side of a stream site or reach. Bankfull width will be measured at the 11 evenly-spaced transects within each reach (for status/trend monitoring) or treatment and control site (for effectiveness monitoring) following the protocol in Bain and Stevenson (1999). Widths of multiple channels are summed to represent the total bankfull width. Sampling will occur once annually during base-flow conditions.

Streambank Condition:

The plan includes streambank stability as the one specific indicator of streambank condition. It will be measured following methods in BURPTAC (1999). Stability will be based on "natural" conditions (e.g., vegetation), not "unnatural" conditions such as car bodies, riprap, and concrete. The methods apply to both the left and right banks of the channel. Bank stability will be measured once annually during base-flow conditions at the 11 evenly-spaced transects within each reach (for status/trend monitoring) or treatment and control site (for effectiveness monitoring).

Riparian Condition

Riparian Habitat:

The plan included percent altered vegetation as the one specific indicator of riparian condition. Percent altered vegetation will be measured within each reach (for status/trend monitoring) or treatment and control site (for effectiveness monitoring) following methods described in Platts et al. (1987). It is measured along both banks. Percent altered vegetation will be measured once annually during base-flow conditions.

Flows and Hydrology

Streamflows:

Changes in streamflows will be assessed by collecting flow data at the downstream end of monitoring reaches and/or at the downstream end of the distribution of each population or subpopulation. Investigators will use USGS flow data where available to assess changes in peak, base, and timing of flows. For those streams with no USGS stream-gauge data, investigators will use methods described in Chapter 14 in Bain and Stevenson (1999) for measuring stream flows. Water velocities will be measured with a calibrated water-velocity meter rather than the float method.

Watershed Conditions

Road Density:

The plan includes road density and the riparian-road index (RRI) as indicators of roads within watersheds. Investigators will measure the road density and riparian-road index within each watershed in which monitoring activities occur. Road density will be calculated with GIS as the total length of roads within a watershed divided by the area of the watershed. The riparian-road index will be calculated with GIS as the total mileage of roads within riparian areas divided by the total number of stream miles within the watershed. WFC (1998) provides an example of calculating the riparian-road index in the Umpqua Basin. Both road density and the riparian-road index will be updated annually.

Watershed Disturbance:

The plan includes "equivalent clearcut area" (ECA) as the single indicator of watershed disturbance. It will be measured within each watershed in which monitoring activities occur. Investigators will follow methods outlined in USFS (1974) and King (1989). ECA will be updated annually.

7.4 Recommendations

This plan requires that the protocols identified above be used to measure biological and physical/environmental indicators. It is understood that some of these methods will differ from those currently used by entities that will be implementing this plan. Indeed, some of the entities that will implement this plan may have collected data for several years with protocols different from those identified in this plan. It is not the intent of this plan to have those entities immediately switch protocols. Rather, this plan encourages entities to use both methods for a few years.[34] This will allow them to compare the performance of different methods and to develop relationships between different protocols.

---

[34] The number of years needed to compare performance and to develop relationships between methods will be determined as data are collected. At a minimum, entities implementing this plan should expect to use both methods for at least five years (n = 5).

SECTION 8: IMPLEMENTATION

The preceding sections serve notice that considerable care must be put into the appropriate methods and logic structure of a status/trend and effectiveness monitoring plan. My intent in this section is to distill the information presented in this document into a concise outline that an investigator can follow to develop a statistically-valid monitoring plan. For convenience, I offer this summary as a checklist of steps that will aid the investigator in developing a valid monitoring program. Although these steps are generic, the investigator should address each one in order to demonstrate complete understanding of status/trend and effectiveness monitoring.

I divided this section into three parts. The first part outlines the steps needed to setup and implement the monitoring plan. The second and third parts outline the steps needed to design status/trend monitoring studies and effectiveness monitoring studies, respectively.

8.1 Program Setup

In order to setup the monitoring program, it is important to follow a logical sequence of steps. By walking through each step, the investigator will better understand the goals of monitoring and its strengths and limitations. These steps should aid the investigator in implementing a valid monitoring program that reduces duplication of sampling efforts, and thus overall costs, but still meets the needs of the different entities. The plan assumes that all entities involved with implementing the plan will cooperate and freely share information.

Setup Steps:

- Identify the populations and/or subpopulations of interest (e.g., spring chinook, steelhead, bull trout).
- Identify the geographic boundaries (areas) of the populations or subpopulations of interest.
- Describe the purpose for selecting these populations or subpopulations (what are the concerns?).
- Identify the objectives for monitoring.
- Select the appropriate monitoring approach (status/trend or effectiveness monitoring or both) for addressing the objectives.
- Identify and review existing monitoring and research programs in the area of interest.
- Determine if those programs satisfy the objectives of the proposed program.
- If data gaps exist, implement the appropriate monitoring approach by following the criteria outlined below.
- Classify the landscape and streams in the area of interest (see Section 5).
- Describe how data collection efforts will be shared among the different entities.
- Identify a common database for storing biological and physical/environmental data.
- Estimate costs of implementing the program.

- Identify cost-sharing opportunities.

8.2 __Status/Trend Monitoring__

If the objective of the monitoring program is to assess the current status of populations and/or environmental conditions, or to assess long-term trends in these parameters, then the following steps will help the investigator design a valid status/trend monitoring program.

Problem Statement and Overarching Issues:

- Identify and describe the problem to be addressed.
- Identify boundaries of the study area.
- Describe the goal or purpose of the study.
- List hypotheses to be tested.

Statistical Design (see Section 2):

- Describe the statistical design to be used (e.g., EMAP design).
- List and describe potential threats to external validity and how these threats will be addressed.
- If this is a pilot test, explain why it is needed.
- Describe descriptive and inferential statistics to be used and how precision of statistical estimates will be calculated.

Sampling Design (see Sections 3 & 4):

- Describe the statistical population(s) to be sampled.
- Define and describe sampling units.
- Identify the number of sampling units that make up the sampling frame.
- Describe how sampling units will be selected (e.g., random, stratified, systematic, etc.).
- Describe variability or estimated variability of the statistical population(s).
- Define Type I and II errors to be used in statistical tests (the plan recommends no less than 0.80 power).

Measurements (see Sections 6 & 7):

- Identify indicator variables to be measured.
- Describe methods and instruments to be used to measure indicators.
- Describe precision of measuring instruments.
- Describe possible effects of measuring instruments on sampling units (e.g., core sampling for sediment may affect local sediment conditions). If such effects are expected, describe how the study will deal with this.
- Describe steps to be taken to minimize systematic errors.
- Describe QA/QC plan, if any.

- Describe sampling frequency for field measurements.

Results:

- Explain how the results of this study will yield information relevant to management decisions.

8.3 Effectiveness Monitoring

If the objective of the monitoring program is to assess the effects of tributary habitat actions (e.g., improve stream complexity), then the following steps will help the investigator design a valid effectiveness monitoring program (these steps are modified from Paulsen et al. 2002). Because effectiveness monitoring encompasses the essence of cause-and-effect research (i.e., attempts to control for sources of invalidity), the steps below are more extensive and intensive than those in the status/trend monitoring program.

Problem Statement and Overarching Issues:

- Identify and describe the problem to be improved or corrected by the action being monitored.
- Describe current environmental conditions at the project site.
- Describe factors contributing to current conditions (e.g., roads crossing causing increased siltation).
- Identify and describe the habitat action(s) (treatments) to be undertaken to improve existing conditions.
- Describe the goal or purpose of the habitat action(s).
- Identify the hypotheses to be tested.
- Identify the independent variables in the study.

Statistical Design (see Section 2):

- Describe the statistical design to be used (e.g., BACI design).
- Describe how treatments (habitat actions) and controls will be assigned to sampling units (e.g., random assignment).
- Show whether or not the study will include “true” replicates or subsamples.
- Describe how temporal and spatial controls will be used and how many of each type will be sampled.
- Describe the independence of treatment and control sites (are control sites completely unaffected by habitat actions?).
- Identify covariates and their importance to the study.
- Describe potential threats to internal and external validity and how these threats will be addressed.
- If this is a pilot test, explain why it is needed.
- Describe descriptive and inferential statistics to be used and how precision of statistical estimates will be calculated.

Sampling Design (see Sections 3 & 4):

- Describe the statistical population(s) to be sampled.
- Define and describe sampling units.
- Describe the number of sampling units (both treatment and control sites) that make up the sampling frame.
- Describe how sampling units will be selected (e.g., random, stratified, systematic, etc.).
- Define "practical significance" (e.g., environmental or biological effects of the action) for this study.
- Describe how effect size(s) will be detected.
- Describe the variability or estimated variability of the statistical population(s).
- Define Type I and II errors to be used in statistical tests (the plan recommends no less than 0.80 power).

Measurements (see Sections 6 & 7):

- Identify and describe the indicator (dependent) variables to be measured.
- Describe methods and instruments to be used to measure indicators.
- Describe the precision of measuring instrument(s).
- Describe possible effects of measuring instruments on sampling units (e.g., core sampling for sediment may affect local sediment conditions). If such effects are expected, describe how the study will deal with this.
- Describe steps to be taken to minimize systematic errors.
- Describe QA/QC plan, if any.
- Describe sampling frequency for field measurements.

Results:

- Explain how the results of this study will yield information relevant to management decisions.

These steps should be considered when designing a monitoring plan to assess the effectiveness of any habitat action, regardless of how simple the proposed action may be. Even monitoring the effectiveness of irrigation screens requires careful consideration of all steps in the checklist. In some cases, the investigator may not be able to address all steps with a high degree of certainty, because adequate information does not exist. For example, the investigator may lack information on population variability, effect size, "practical significance," or instrument precision, which makes it difficult to design studies and estimate sample sizes. In this case the investigator can address the statements with the best available information, even if it is based on professional opinion, or design a pilot study to answer the questions.

SECTION 9: APPLICATION

SECTION 10: REFERENCES

Action Agencies (Bonneville Power Administration, U.S. Bureau of Reclamation, and U.S. Army Corps of Engineers) and NOAA Fisheries. 2003. Draft research, monitoring and evaluation plan for the National Marine Fisheries Service 2000 Federal Columbia River Power System biological opinion. Bonneville Power Administration, U.S. Bureau of Reclamation, U.S. Army Corps of Engineers, and NOAA Fisheries, Portland, OR.

Anderson, R. O. and R. M. Neumann. 1996. Length, weight, and associated structural indices. Pages 447-482 *in*: B. R. Murphy and D. W. Willis, editors. Fisheries techniques, 2nd edition. American Fisheries Society, Bethesda, MD.

Armantrout, N. B., compiler. 1998. Glossary of aquatic habitat inventory terminology. American Fisheries Society, Bethesda, MD.

Bain, M. B. and N. J. Stevenson, editors. 1999. Aquatic habitat assessment: common methods. American Fisheries Society, Bethesda, MD.

Bailey, R. G. 1978. Description of ecoregions of the United States. U.S. Forest Service, Intermountain Region, Ogden, UT.

Bailey, R. G. 1998. Ecoregions map of North America: explanatory note. U.S. Forest Service, Miscellaneous Publication 1548, Washington, D.C.

Bevenger, G. S. and R. M. King. 1995. A pebble count procedure for assessing watershed cumulative effects. Research Paper RM-RP-319, USDA Forest Service, Rocky Mountain Forest and Range Experiment Station, Fort Collins, CO.

Bilhimer, D., J. Carroll, S. O'Neal, and G. Pelletier. 2003. Draft quality assurance project plan: Wenatchee River temperature, dissolved oxygen, pH, and fecal coliform total maximum daily load year 2 technical study. Washington State Department of Ecology, Olympia, WA.

Bisson, P. A. and D. R. Montgomery. 1996. Valley segments, stream reaches, and channel units. Pages 23-52 *in*: R. R. Hauer and G. A. Lamberti, editors. Methods in stream ecology. Academic Press, New York, NY.

Bjornn T. C. and D. W. Reiser. 1991. Habitat requirements of salmonids in streams. American Fisheries Society Special Publication 19:83-138.

Borgerson, L. A. 1992. Scale analysis. Oregon Department of Fish and Wildlife, Fish Research Project F-144-R-4, Annual Progress Report, Portland, OR.

Box, G. E. P. and G. M. Jenkins. 1976. Time-series analysis: forecasting and control. Holden-Day, San Francisco, CA.

Browne, R. H. 2001. Using the sample range as a basis for calculating sample size in power calculations. The American Statistician 55:293-298.

Bunte, K. and S. R. Abt. 2001. Sampling frame for improved pebble count accuracy in coarse gravel-bed streams. Journal of the American Water Resources Association 37:1001-1014.

BURPTAC (Beneficial Use Reconnaissance Project Technical Advisory Committee). 1999. 1999 beneficial use reconnaissance project workplan for wadable streams. Idaho Division of Environmental Quality, Boise, ID.

Cailliet, G. M., M. S. Love, and A. W. Ebeling. 1986. Fishes, a field and laboratory manual on their structure, identification, and natural history. Wadsworth Publishing Company, Belmont, CA.

Clay, C. H. 1995. Design of fishways and other fish facilities. Second edition. Lewis Publishers, Boca Raton, Florida.

Cohen, J. 1988. Statistical power analysis for the behavioral sciences. Lawrence-Erlbaum, Hillsdale, NJ.

Cook, T. D. and D. T. Campbell. 1979. Quasi-experimentation: design and analysis issues for field settings. Houghton Mifflin Company, Boston, MA.

Cupp, C. E. 1989a. Identifying spatial variability of stream characteristics through classification. Master's thesis. University of Washington, Seattle, WA.

Cupp, C. E. 1989b. Valley segment type classification for forested lands of Washington. Washington State Timber/Fish/Wildlife Agreement, TFW-AM-89-001, Department of Natural Resources, Olympia, WA.

Diaz-Ramos, S., D. L. Stevens, and A. R. Olsen. 1996. EMAP statistical methods manual. U.S. Environmental Protection Agency, EPA/620/R-96/XXX, Corvallis, OR.

Dolloff, A., J. Kershner, and R. Thurow. 1996. Underwater observation. Pages 533-554 *in*: B. R. Murphy and D. W. Willis, editors. Fisheries techniques, 2nd edition. American Fisheries Society, Bethesda, MD.

Frissell, C. A., W. J. Liss, C. E. Warren, and M. D. Hurley. 1986. A hierarchical framework for stream habitat classification; viewing streams in a watershed context. Environmental Management 10:199-214.

Green, R. H. 1979. Sampling design and statistical methods for environmental biologists. John Wiley and Sons, Inc., New York, NY.

Green, R. H. 1994. Aspects of power analysis in environmental monitoring. Pages 173-182 *in*: D. J. Fletcher and B. F. J. Manly, editors. Statistics in ecology and environmental monitoring. University of Otago Press, Dunedin.

Gregory, S. V. and P. A Bisson. 1997. Degradation and loss of anadromous salmonid habitat in the Pacific Northwest. Pages 277-314 in: D. J. Stouder, P. A. Bisson, and R. J. Naiman, editors. Pacific salmon and their ecosystems, status and future options. Chapman and Hall, New York, NY.

Hadley, R. F. and S. A. Schumm. 1961. Sediment sources and drainage basin characteristics in upper Cheyenne River basin. U.S. Geological Survey, Water-Supply Paper 1531-B, Reston, Virginia.

Hairston, N. G. 1989. Ecological experiments: purpose, design, and execution. Cambridge University Press, New York, NY.

Hicks, L. L., J. Light, G. Watson, B. Sugden, T. W. Hillman, and D. Berg. 1999. Adaptive management: concepts and applications to Plum Creek's Native Fish Habitat Conservation Plan. Native Fish Habitat Conservation Plan Technical Report No. 13, Plum Creek Timber Company, Seattle, WA.

Hillman, T. W. and D. W. Chapman. 1996. Comparison of underwater methods and electrofishing for estimating fish populations in the Upper Blackfoot River Basin. BioAnalysts, Inc. Report to the Seven-Up Pete Joint Venture, Lincoln, MT.

Hillman, T. W. and A. E. Giorgi. 2002. Monitoring protocols: effectiveness monitoring of physical/environmental indicators in tributary habitats. BioAnalysts, Inc. Report to Bonneville Power Administration, Portland, Or. http://www.efw.bpa.gov/cgi-bin/FW/welcome.cgi?ViewMode=External

Hillman, T. W. and M. D. Miller. 2002. Abundance and total numbers of chinook salmon and trout in the Chiwawa River Basin, Washington. BioAnalysts, Inc. Report to Chelan County Public Utility District, Wenatchee, WA.

Hughes, R. M., E. Rexstad, and C. E. Bond. 1987. The relationship of aquatic ecoregions, river basins and physiographic provinces to the ichthyogeographic regions of Oregon. Copeia 2:423-432.

Hunt, C. B. 1967. Physiography of the United States. W. H. Freeman, San Francisco, CA.

Hurlbert, S. J. 1984. Pseudoreplication and the design of ecological field experiments. Ecological Monographs 54:187-211.

Idea Works, Inc. 1997. Methodologists toolchest. Version 1.2. Distributed by Scolari, Sage Publications Software, Beverly Hills, CA.

ISAB (Independent Scientific Advisory Board). 2003. A review of strategies for recovering tributary habitat. Independent Scientific Advisory Board for the Northwest Power Planning Council, ISAB 2003-2, Portland, OR.

Johnson, D. H., and nine others. 2001. Inventory and monitoring of salmon habitat in the Pacific Northwest—directory and synthesis of protocols for management/research and volunteers in Washington, Oregon, Montana, Idaho, and British Columbia. Review draft. Washington Department of Fish and Wildlife, Olympia, WA.

King, J. G. 1989. Streamflow responses to road building and harvesting: a comparison with the equivalent clearcut area procedure. USDA Forest Service, Research Paper INT-401, Ogden, UT.

Krebs, C. J. 1999. Ecological methodology. Second edition. Benjamin/Cummings, Menlo Park, CA.

Lazorchak, J. M., D. J. Klemm, and D. V. Peck (editors). 1998. Environmental monitoring and assessment program—surface waters: field operations and methods for measuring the ecological condition of wadeable streams. EPA/620/R-94/004F, U.S. Environmental Protection Agency, Washington, D.C.

Lee, K. N. 1993. Compass and gyroscope: integrating science and politics for the environment. Island Press, Washington, D.C.

Lipsey, M. W. 1990. Design sensitivity: statistical power for experimental research. Sage Publications, Beverly Hills, CA.

MacDonald, L. H., A. W. Smart, and R. C. Wissmar. 1991. Monitoring guidelines to evaluate effects of forestry activities on streams in the Pacific Northwest and Alaska. U.S. Environmental Protection Agency, EPA/910/9-91-001, Seattle, WA.

Manly, B. F. J. 1992. The design and analysis of research studies. Cambridge University Press, New York, NY.

Manly, B. F. J. 2001. Statistics for environmental science and management. Chapman and Hall, New York, NY.

Mapstone, B. D. 1995. Scalable decision rules for environmental impact studies: effect size, Type I, and Type II errors. Ecological Applications 5:401-410.

Meehan, W. R., editor. 1991. Influences of forest and range management on salmonid fishes and their habitats. American Fisheries Society Special Publication 19, Bethesda, MD.

Montgomery, D. R. and J. M. Buffington. 1993. Channel classification, prediction of channel response, and assessment of channel condition. Washington State Timber/Fish/Wildlife Agreement, TFW-SH10-93-002, Department of Natural Resources, Olympia, WA.

Moore, K. 2002. Draft Oregon plan for salmon and watersheds monitoring strategy. Oregon Plan, Salem, OR.

Mosey, T. R. and L. J. Murphy. 2002. Spring and summer chinook spawning ground surveys on the Wenatchee River Basin, 2001. Chelan County Public Utility District, Wenatchee, WA.

Murdoch, A., K. Petersen, T. Miller, M. Tonseth, and T. Randolph. 1999. Freshwater production and emigration of juvenile spring chinook salmon from the Chiwawa River in 1998. Report No. SS99-05, Washington Department of Fish and Wildlife, Olympia, WA.

Naiman, R. J., D. G. Lonzarich, T. J. Beechie, and S. C. Ralph. 1992. General principles of classification and the assessment of conservation potential in rivers. Pages 93-123 *in*: P. J. Boon, P. Calow, and G. E. Petts, editors. River conservation and management. John Wiley and Sons, New York, NY.

Nawa, R. K., C. A. Frissell, and W. J. Liss. 1988. Life history and persistence of anadromous fish stocks in relation to stream habitats and watershed classification. Annual progress report to Oregon Department of Fish and Wildlife, Portland, OR.

NCSS (Number Cruncher Statistical Systems). 2000. Pass 2000 power analysis and sample size for Windows. NCSS, Kaysville, UT.

NMFS (National Marine Fisheries Service). 1996. Making Endangered Species Act determinations of effect for individual or grouped actions at the watershed scale. The National Marine Fisheries Service Environmental and Technical Services Division, Habitat Conservation Branch, Seattle, WA.

NRC (National Research Council). 1992. Restoration of aquatic ecosystems: science, technology, and public policy. National Academy Press, Washington, D.C.

Omernik, J. M. 1987. Aquatic ecoregions of the conterminous United States. Annals of the Association of American Geographers 77:118-125.

OPSW (Oregon Plan for Salmon and Watersheds). 1999. Water quality monitoring, technical guide book. Version 2.0. Corvallis, OR.

Overton, C. K., S. P. Wollrab, B. C. Roberts, and M. A. Radko. 1997. R1/R4 (Northern/Intermountain Regions) fish and fish habitat standard inventory procedures handbook. USDA Forest Service General Technical Report INT-GTR-346, Ogden, UT.

Overton, W. S., D. White, and D. L. Stevens. 1990. Design report for EMAP environmental monitoring and assessment program. U.S. Environmental Protection Agency, EPA/600/3-91/053, Corvallis, OR.

Parker, M. A. 2000. Fish passage – culvert inspection procedures. Watershed Restoration Technical Circular No. 11. Ministry of Environment, Lands and Parks and Ministry of Forest, British Columbia.

Paulsen, C., S. Katz , T. Hillman, A. Giorgi, C. Jordan, M. Newsom, and J. Geiselman, 2002 Guidelines for Action Effectiveness Research Proposals for FCRPS Offsite Mitigation Habitat Measures. Bonneville Power Administration, U.S. Bureau of Reclamation, U.S. Army Corps of Engineers, and NOAA Fisheries, Portland, OR. http://www.efw.bpa.gov/cgiin/FW/welcome.cgi?ViewMode=ExternalView.

Peterman, R. M. 1990. Statistical power analysis can improve fisheries research and management. Canadian Journal of Fisheries and Aquatic Sciences 47:2-15.

Platts, W. S., W. F. Megahan, and G. W. Minshall. 1983. Methods for evaluating stream, riparian, and biotic conditions. USDA Forest Service General Technical Report INT-138, Ogden, UT.

Platts, W. S. and twelve others. 1987. Methods for evaluating riparian habitats with applications to management. USDA Forest Service General Technical Report INT-221, Ogden, UT.

Reeves, G. H., and nine others. 2001. Aquatic and riparian effectiveness monitoring plan for the Northwest Forest Plan. USDA Forest Service, Pacific Northwest Research Station, Corvallis, OR.

Reynolds, J. B. 1996. Electrofishing. Pages 221-253 *in*: B. R. Murphy and D. W. Willis, editors. Fisheries techniques, 2nd edition. American Fisheries Society, Bethesda, MD.

Rosgen, D. 1996. Applied river morphology. Wildland Hydrology, Pagosa Springs, CO.

Scheaffer, R. L., W. Mendenhall, and L. Ott. 1990. Elementary survey sampling. Fourth edition. PWS-KENT Publishing Company, Boston, MA.

Schuett-Hames, D., J. Ward, M. Fox, A. Pleus, and J. Light. 1994. Large woody debris survey module. Section 5 *in*: D. Schuett-Hames, A. Pleus, L. Bullchild, and S. Hall, editors. Ambient monitoring program manual. Timber-Fish-Wildlife TFW-AM9-94-001, Northwest Indian Fisheries Commission, Olympia, WA.

Schuett-Hames, D., A. E. Pleus, E. Rashin, and J. Matthews. 1999a. Method manual for the stream temperature survey. Timber-Fish-Wildlife TFW-AM9-99-005, Northwest Indian Fisheries Commission, Olympia, WA.

Schuett-Hames, D., R. Conrad, A. Pleus, and M. McHenry. 1999b. Method manual for the salmonid spawning gravel composition survey. Timber-Fish-Wildlife TFW-AM9-99-006, Northwest Indian Fisheries Commission, Olympia, WA.

Simonson, T., J. Lyons, and P. Kanehl. 1994. Quantifying fish habitat in streams: transect spacing, sample size, and a proposed framework. North American Journal of Fisheries Management 14:607-615.

Skalski, J. R. and D. S. Robson. 1992. Techniques for wildlife investigations, design and analysis of capture data. Academic Press, New York, NY.

Smith, E. P., D. R. Orvos, and J. Cairns. 1993. Impact assessment using the before-after-control-impact (BACI) model: concerns and comments. Canadian Journal of Fisheries and Aquatic Sciences 50:627-637.

Spence, B. C., G. A. Lomnicky, R. M. Hughes, and R. P. Novitzki. 1996. An ecosystem approach to salmonid conservation. TR-4501-96-6057, Management Technology, Corvallis, OR.

Stevens, D. L. 1997. Variable density grid-based sampling designs for continuous spatial populations. Environmetrics 8:167-195.

Stevens, D. L. 2002. Sampling design and statistical analysis methods for the integrated biolgocial and physical monitoring of Oregon streams. Report No. OPSW-ODFW-2002-07, The Oregon Plan for Salmon and Watersheds, Oregon Department of Fish and Wildlife, Corvallis, OR.

Stevens, D. L. and A. R. Olsen. 1999. Spatially restricted surveys over time for aquatic resources. Journal of Agricultural, Biological, and Environmental Statistics 4:415-428.

Stevens, D. L. and N. S. Urquhart. 2000. Response designs and support regions in sampling continuous domains. Environmetrics 11:13-41.

Stewart-Oaten, A., W. W. Murdoch, and K. R. Parker. 1986. Environmental impact assessment: “pseudoreplication” in time? Ecology 67:929-940.

Stewart-Oaten, A., J. R. Bence, and C. W. Osenberg. 1992. Assessing effects of unreplicated perturbations: no simple solutions. Ecology 73:1396-1404.

Strange, R. J. 1996. Field examination of fishes. Pages 433-446 *in*: B. R. Murphy and D. W. Willis, editors. Fisheries techniques, 2nd edition. American Fisheries Society, Bethesda, MD.

Underwood, A. J. 1994. Things environmental scientists (and statisticians) need to know to receive (and give) better statistical advice. Pages 33-61 *in*: D. J. Fletcher and B. F. J. Manly, editors. Statistics in ecology and environmental monitoring. University of Otago Press, Dunedin.

USFS (U.S. Forest Service). 1974. Forest hydrology part II—hydrologic effects of vegetation manipulation. USDA Forest Service, Missoula, MT.

USFWS (U.S. Fish and Wildlife Service). 2000. Biological Opinion, effects to listed species from operations of the Federal Columbia River Power System. U.S. Fish and Wildlife Service, Regions 1 and 6, Portland, OR.

Van Deventer, J. S. and W. S. Platts. 1989. Microcomputer software system for generating population statistics from electrofishing data—user's guide for MicroFish 3.0. USDA Forest Service General Technical Report INT-254, Ogden, UT.

WDFW (Washington Department of Fish and Wildlife). 2000. Fish passage barrier and surface water diversion screening assessment and prioritization manual. Washington Department of Fish and Wildlife Habitat Program, Environmental Restoration Division, Olympia, WA.

Wertz, W. A. and J. F. Arnold. 1972. Land systems inventory. U.S. Forest Service, Intermountain Region, Ogden, UT.

WFC (World Forestry Center). 1998. Pilot study report, Umpqua land exchange project. World Forestry Center, Portland, OR.

WFPB (Washington Forest Practices Board). 1995. Washington forest practices board manual: standard methodology for conducting watershed analysis under Chapter 222-22 WAC. Version 3.0. Washington Forest Practices Board, Olympia, WA.

Whittier, T. R., R. M. Hughes, and D. P. Larsen. 1988. Correspondence between ecoregions and spatial patterns in stream ecosystems in Oregon. Canadian Journal of Fisheries and Aquatic Sciences 45:1264-1278.

Wipfli, M. S. and D. P. Gregovich. 2002. Export of invertebrates and detritus from fishless headwater streams in southeastern Alaska: implications for downstream salmonid production. Freshwater Biology 47:957-969.

WSRFB (Washington Salmon Recovery Funding Board). 2003. Draft 5/16/2003 monitoring and evaluation strategy for habitat restoration and acquisition projects. Washington Salmon Recovery Funding Board, Olympia, WA.

Zaroban, D. W. 2000. Protocol for placement and retrieval of temperature data loggers in Idaho streams. Idaho Division of Environmental Quality, Boise, ID.